Individual Differences and Instructed Second Language Acquisition: Insights from Intelligent Computer Assisted Language Learning

D i s s e r t a t i o n
zur
Erlangung des akademischen Grades
Doktor der Philosophie
in der Philosophischen Fakultät

der Eberhard Karls Universität Tübingen

vorgelegt von

Simón Eduardo Ruiz Hernández

aus

Caracas, Venezuela

2019

Gedruckt mit Genehmigung der Philosophischen Fakultät der
Eberhard Karls Universität Tübingen

Dekan: Prof. Dr. Jürgen Leonhardt

Hauptberichterstatter: Prof. Dr. Detmar Meurers
Mitberichterstatter: Prof. Dr. Barbara Kaup

Tag der mündlichen Prüfung: 21. Dezember 2018

Universitätsbibliothek Tübingen, TOBIAS-lib

Para mis papás, Simón y María Auxiliadora

Para mi familia

To Rebecca Beke, in memoriam

Acknowledgements

First and foremost, I would like to thank my main supervisor, Professor Detmar Meurers, for his generous and constant support and guidance. I have cherished your tutelage, and I have learned so much throughout these years. Thank you very much for introducing me to the world of LaTex. It was/is painful at times, but worthy all the way! I would also like to thank my co-supervisor, Professor Barbara Kaupp, for her valuable advice and suggestions.

I am also in debt with Patrick Rebuschat, who has provided me with unflagging support. I feel very lucky to have crossed paths with Patrick, and I appreciate all of the encouragement and help he has given me since day one.

My special thanks to my colleagues and friends Xiaobin Chen, Katharina Wendebourg, Björn Rudzewitz, Martí Quixal, and Maria Chinkina, who have always been there for me when I needed them. Your academic and personal support certainly contributed to the completion of this work.

I also wish to express my deepest gratitude to the LEAD Graduate School and Research Network not only for financially supporting this dissertation, but also for providing me with unparallel assistance during my doctoral studies. I am specially grateful to Mareike Bierlich, Sophie Freitag, and Johann Jacoby. In particular, I would like to thank Johann, for guiding and helping me analyze the data using mixed-effects regression modeling. I very much enjoyed our discussions and learned a lot in the process.

I would like to thank our research assistants Aria Omidvar, Stephanie Dawn

Samson and Peter Schoener, who were instrumental in the technical aspects of the dissertation.

I owe thanks as well to the administrative staff in the Department of Linguistics at the University of Tübingen, especially Lisa Becker.

I am also thankful to Kara Morgan-Short, Sarah Grey, and Kaitlyn Tagarelli for their feedback and support throughout this research project.

I am endlessly grateful as well to all the people who trusted in me over the years. I would like to single out and thank Yuly Asencion Delaney, Rebecca Beke, Mirna Quintero, Rosynella Cardozo, María Antonini, Paola Bentivoglio, Carmen Mendoza, Juan Pino-Silva, Adriana Gónzalez, Víctor Ojeda, Wilfredo Sequera, Miguel Mendoza, and Carlos Mayora. Their support and help paved the way to the achievement of this academic goal.

Last but not least, my heartfelt appreciation goes out to María José, *my compañera de aventuras*. It was not easy but your countless sacrifices, inconditional love, and never-ending patience kept me going. ¡Gracias!

Abstract

The present dissertation focuses on the role of cognitive individual difference factors in the acquisition of second language vocabulary in the context of intelligent computer assisted language learning (ICALL). The aim was to examine the association between working memory and declarative memory and the learning of English phrasal verbs in a web-based ICALL-mediated experiment. Following a pretest-posttest design, 127 adult learners of English were assigned to two instructional conditions, namely meaning-focused and form-focused conditions. Learners in both conditions read news texts on the web for about two weeks; learners in the form-focused condition additionally interacted with the texts via selecting multiple-choice options.

The results showed that both working memory and declarative memory were predictive of vocabulary acquisition. However, only the working memory effect was modulated by the instructional context, with the effect being found exclusively in the form-focused condition, and thus suggesting the presence of an aptitude-treatment interaction. Finally, findings also revealed that learning during treatment in the form-focused group was nonlinear, and that paying attention to form and meaning simultaneously impeded global reading comprehension for intermediate, not advanced learners.

From a theoretical perspective, the findings provide evidence to suggest that individual differences in both working memory and declarative memory affect the acquisition of lexical knowledge in ICALL-supported contexts. Methodologically, the current study illustrates the advantages of conducting interdisciplinary work between ICALL and second language acquisition by allowing for the collection of experimental data through a

web-based, all-encompassing ICALL system. Overall, the present dissertation represents an initial attempt at characterizing who is likely to benefit from ICALL-based interventions.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

In the field of second language acquisition (SLA), there is a general consensus that instruction is beneficial (Kang, Sok, & Han, 2018; Loewen & Sato, 2017; Norris & Ortega, 2000). For instance, instruction may increase learning rate and push ultimate attainment. As such, instruction has been generally grouped into two major types, namely meaning-focused and form-focused instruction (Ellis, 2001, 2012; Norris & Ortega, 2000). The former type refers to the instruction whereby learners are required to attend only to the content of communication and learning is a by-product of both exposure to rich and comprehensible input, as well as communication (Ellis, 2001; Krashen & Terrell, 1983). As mere exposure to input may be insufficient for acquisition to occur (Long, 1991; Swain, 1995), the latter type of instruction seeks to draw learners' attention to linguistic form, in hope that increased attention to target forms will facilitate learning (Boers, 2017; Loewen, 2011; Schmidt, 1990; Sharwood Smith, 1991, 2013). The benefits of instruction may, however, depend on learners' individual differences, such as their cognitive ability to store and process information simultaneously (Baddeley, 2017).

Individual differences are regarded as critical factors in acquiring and processing language (Kidd, Donnelly, & Christiansen, 2018) and they influence the effects of instruction (Snow, 1991). In second language (L2) learning, they have been shown to have great explanatory power in accounting for L2 learning outcomes (Dörnyei, 2005;

Ellis, 2012). In fact, individual differences are indicated as the most significant predictors of L2 learners' success (Dörnyei, 2005; Kormos, 2013). In the present study, the focus is on two very important individual difference variables, namely working memory and declarative memory. These cognitive factors are deemed crucial sources of learner variation (Buffington & Morgan-Short, in press; Hamrick, Lum, & Ullman, 2018; Li, 2017a; Tagarelli, Borges Mota, & Rebuschat, 2015; Tagarelli, Ruiz, Moreno Vega, & Rebuschat, 2016).

Recent L2 research has increasingly documented that individual differences interact with instructional conditions or *treatments* (e.g., Erlam, 2005; Faretta-Stutenberg & Morgan-Short, 2018; Li & Fu, 2016; Malone, 2018; Suzuki & DeKeyser, 2017). This kind of research examines the extent to which pedagogical treatments are more or less effective depending on learners' cognitive abilities or *aptitudes*, such as working memory, declarative memory or procedural memory (Buffington & Morgan-Short, in press; Vatz, Tare, Jackson, & Doughty, 2013). Examining these aptitude-treatment interactions (Cronbach & Snow, 1977) will potentially inform the design of pedagogical interventions that can suit learners' characteristics, thereby optimizing L2 learning (Li, 2017a; Suzuki & DeKeyser, 2017). Crucially, aptitude-treatment interaction research may also provide insights into the nature of L2 learning processes by ascertaining the potential facilitative or hampering role of certain cognitive abilities in the success or failure of specific instructional situations (DeKeyser, 2012b; Robinson, 2002; Suzuki & DeKeyser, 2017).

In the current dissertation, this important line of inquiry is expanded by investigating the role of individual differences on L2 vocabulary acquisition and its potential interaction with instructional conditions in a web-based learning environment. More specifically, it examined i) the influence of working memory and declarative memory on the learning of English phrasal verbs under meaning-focused and form-focused conditions, as well as ii) possible aptitude-treatment interactions. With this in mind, a web-based intelligent computer-assisted language learning (ICALL) system was used to conduct the current study.

ICALL can be understood as the use of natural language processing (NLP) technology in computer-assisted language learning (CALL). NLP is considered to be the applied side of computational linguistics that supports a potentially rich interface with language teaching and learning (Lu, 2018; Meurers, 2013; Meurers & Dickinson, 2017). As such, ICALL systems can be used to gain insights into the processes and products of L2 acquisition. Moreover, ICALL systems offer opportunities for large scale, systematic administrations of testing and treatments (Meurers & Dickinson, 2017; Ziegler et al., 2017). In the present study, a web-based ICALL platform, an adapted version of the WERTi system (Meurers et al., 2010), was employed to carry out an entire experiment on the web. Collecting empirical data via the Internet can significantly enhance the study of SLA, as it facilitates the collection of larger amounts of data in shorter time periods than is typically feasible in lab-based experiments (MacWhinney, 2017; Meurers & Dickinson, 2017).

The present dissertation is organized as follows. Chapter 2 reviews the relevant literature on instructed SLA, individual differences, and ICALL. Chapter 3 describes the selection of the target linguistic form. This chapter is further divided into two parts. The first part provides an overview of the literature on the acquisition of English phrasal verbs, and the second reports the results of a corpus-based[1] study that allowed the selection of learning targets and materials. Chapter 4 reports the results of a study that compared performance on the lab-based and web-based versions of the cognitive tests used here. Chapter 5 details the methodology of the study. Chapter 6 presents the results of the data analysis. Chapter 7 discusses the results of the data analysis in light of the research questions. Chapter 8 provides an overall discussion of the findings, discusses the limitations of the study, suggests directions for future research, and considers the pedagogical implications of the findings. And finally, Chapter 9 summarizes the main findings and draws some conclusions.

[1] A corpus is a large electronic collection of texts (Biber, Conrad, & Reppen, 1998; Gardner & Davies, 2007).

Part I

Chapter 2

Literature review

The current chapter consists of four sections. The first reviews research on the area of instructed SLA. The next discusses empirical work investigating the association of individual differences and L2 development, with a focus on the interaction of cognitive individual differences (e.g., working memory and declarative memory) and instructional procedures, i.e., aptitude-treatment interaction). The third section considers the conceptualization of ICALL and its potential. In this section, recent research on the area will also be reviewed. Finally, the chapter presents the purpose and the research questions addressed in this dissertation.

2.1 Instructed Second Language Acquisition

There is general agreement among SLA researchers that *instruction*, understood here as "any [systematic,] deliberate attempt to promote language learning by manipulating the mechanisms of learning and/or the conditions under which these operate" (de Graaff & Housen, 2009, p. 726), is beneficial for the acquisition of a second language (Benati, 2016; de Graaff & Housen, 2009; Doughty, 2003; Ellis, 2001, 2008, 2012; Goo, Granena, Yilmaz, & Novella, 2015; Housen & Pierrard, 2005; Loewen & Sato, 2017; Long, 1983, 1991; Nassaji, 2016; Norris & Ortega, 2000; Schmidt, 1990; Spada, 1997, 2011; Spada & Tomita, 2010; VanPatten & Benati, 2015).

Instruction can positively affect L2 learning in any of the following ways, among others: i) it may increase acquisition rate (i.e., instructed learners learn faster than uninstructed learners); ii) it may push ultimate attainment (i.e., instructed learners may attain higher proficiency levels than uninstructed learners); iii) it may alter the order in which the acquisition of target forms (i.e., formal aspects of the language such as grammar) occurs; iv) it can provide learners with input and output opportunities (i.e., exposure) to the target language (i.e., the language learners are attempting to learn) that may be otherwise unavailable; iv) it can boost learners' motivation to learn the target language; v) it can help consolidate L2 knowledge so that learners are able to perform a wider range of tasks with greater ease (i.e., they become more fluent); vi) it can assist the acquisition of new L2 forms (i.e., knowledge internalization), causing learners' language to become more elaborate; vii) it can facilitate the modification or restructuring of acquired knowledge, allowing L2 learners to produce more accurate language; and it can facilitate the *noticing* of linguistic forms (Doughty, 2003; Ellis, 1994, 2016; Housen & Pierrard, 2005; Klein, 1986; Schmidt, 1990; Sharwood Smith, 1991, 1993, 2013; Spada & Lightbown, 1999).

2.1.1 Types of L2 instruction

Instruction in L2 learning has been classified into two broad types, meaning-focused instruction and form-focused instruction, depending on whether the emphasis is on meaning or linguistic form, respectively (de Graaff & Housen, 2009; Ellis, 2001, 2008; Housen & Pierrard, 2005; Loewen, 2011; Loewen & Sato, 2017; Norris & Ortega, 2000; Spada, 1997, 2011). Meaning-focused instruction concerns the instruction in which learners are asked to attend only to the content of communicative-driven activities, such as reading texts for meaning (de Graaff & Housen, 2009; Ellis, 2001, 2008; Housen & Pierrard, 2005; Krashen, 1981, 1985; Laufer, 2005; Loewen, 2011; Loewen & Sato, 2017; Norris & Ortega, 2000; Spada, 1997, 2011). In this view, language is a tool for communication, and, therefore, the overall purpose of L2 instruction should be on the communication of meanings (Loewen, 2011). Meaning-focused instruction is based on the premise that learners can acquire the target language incidentally, that is, without intention, as a by-product of communicative activities, and by being exposed to rich and comprehensible L2 input (Boers, 2017; Housen & Pierrard, 2005; Krashen, 1981, 1985; Loewen & Sato, 2017; Long, 1991; Long & Robinson, 1998). Learning discreet linguistic items and grammar rules is thus considered unnecessary and unfruitful (Krashen, 1981, 1985; Schwartz, 1993). Consequently, in meaning-focused instruction, there is generally no direct, overt attention to language forms. In this sense, meaning-focused instruction can also be characterized as implicit (see below). Examples of this kind of instruction are communicative language teaching (Littlewood, 2011; Loumbourdi, 2018), task-based language teaching (Ellis, 2017; Moore, 2018), and content-based instruction (Lyster, 2017; Waller, 2018). It is worth noting that, even when the emphasis is on communication, some of these types of meaning-focused instruction may also encourage some attention to linguistic forms (Loewen, 2011).

Form-focused instruction, on the other hand, pertains to the instruction in which learners' attention is drawn to linguistic form in an attempt to assist learning (e.g., rule explanation; de Graaff and Housen 2009; Ellis 2001, 2008, 2016; Long 1991; Robinson, Mackey, Gass, and Schmidt 2012; Schmidt 1990, 2001, 2010; Sharwood Smith 1991, 1993, 2013). In this case, language is an object to be learned (Ellis, 2001). Form-focused instruction is premised on the idea that, although extensive and meaningful exposure to *input* (i.e., the language that is available in the learner's environment; Zyzik, 2012) is essential to successfully acquire a foreign language (Ellis, 2015a; Leow, 2015; Robinson et al., 2012), mere exposure to input is insufficient for L2 acquisition to occur (Long, 1991; Swain, 1985, 1995; Wong, 2004). Therefore, and based on the Noticing Hypothesis (Schmidt, 1990, 2001, 2010) that states that paying conscious attention to the forms/structures of the targeted language (i.e., *noticing*) is a precondition for learning, it is argued that learners need to *notice* certain features of the language (e.g., grammatical structures, lexical items, phonological features) in the input in order for these features to be internalized (Housen & Pierrard, 2005; Robinson et al., 2012; Schmidt, 1990, 2001, 2010; Sharwood Smith, 1991, 1993, 2013). In other words, to learn the new language, learners need to pay attention to its categories and forms. Moreover, by drawing learners' attention to formal aspect of the L2 language, form-focused instruction renders these aspects more salient, which in turn may help learners become aware (conscious) of certain language features that would otherwise go unnoticed (Housen & Pierrard, 2005; Robinson et al., 2012; Schmidt, 2001, 2010; Sharwood Smith, 1991, 1993, 2013).

2.1.2 Types of form-focused instruction

Form-focused instruction can be of two main kinds, one consisting of deliberate attempts to intervene the process of L2 development, and one consisting of attempts to induce attention to linguistic form during instruction with no explicit aim to teach them (Ellis, 2001, 2012, 2015a). These two types of form-focused instruction have been associated with Long's (1988; 1991) terms *focus on forms* and *focus on form*, respectively. The former invites *intentional learning*, i.e., there is a cognitive effort on the part of the learner to commit target language forms to memory, whereas the latter invites *incidental learning*, i.e., there is an absence of intention to learn target language forms in the L2 input (Ellis, 2015a; Hulstijn, 2003, 2005; Leow & Zamora, 2017).

Focus on forms has been described as a pedagogic approach that consists of teaching discreet linguistic forms (Ellis, 2015a; Loewen, 2011; Long, 1991). In this approach, i) the target form is pre-selected before instruction, ii) learners know what target form they are supposed to be learning, and iii) there are opportunities for intensive exposure and practice of the target form (Ellis, 2015a; Long, 1991). By contrast, focus on form has traditionally been seen as a pedagogic approach whose aim is to draw learners' attention to linguistic items as they arise incidentally while the focus remains on meaning or communication (Ellis, 2015a; Long, 1991). In this approach, i) interaction is encouraged, ii) meaning is central, and iii) attention to target forms is incidental (Ellis, 2015a;

Long, 1991).

Although both focus on forms and focus on form have been considered opposing instructional approaches, Ellis (2016; see also Ellis, 2015b) recently argues that this conceptualization needs to be revised. Ellis proposes that they be treated not as approaches, but, instead, they are to be regarded as different types of instructional procedures. The justification for this is given by the fact that attention to linguistic form has a significant role in both approaches. In this manner, focus on forms then refers to different instructional devices (e.g., exercises) that are by design intended to direct learners' attention to specific linguistic forms, which are to be learned as objects. Focus on form, on the other hand, includes a set procedures for attracting learners' attention to linguistic forms as learners employ language for communication. By treating focus on form as procedures, Ellis (2016) contends, also avoids the issue of whether it can be applied incidentally (i.e., no pre-selection of target forms) or whether it can be set up in advance.

2.1.3 Explicit and implicit instruction

Form-focused instruction has been further categorized according to the explicitness of the instruction (de Graaff & Housen, 2009; Ellis, 2001, 2012; Goo et al., 2015; Nassaji, 2016; Norris & Ortega, 2000; Spada, 1997, 2011; Spada & Tomita, 2010). In this sense, instruction can range in a continuum from *implicit* to *explicit*. This categorization is grounded on psychological considerations of whether instruction is designed to foster learning with or without awareness, i.e., learners are conscious or not of what they are learning, respectively (Ellis, 2012). In explicit instruction, learners are presented with information about rules or are directly instructed to pay attention to particular forms. In implicit instruction, however, no rule presentation or instructions to pay attention to any particular forms are given (DeKeyser, 1995; Ellis, 1994; Goo et al., 2015; Hulstijn, 2005; Norris & Ortega, 2000).

Explicit instruction has two main purposes, namely i) to direct learners' attention to forms and ii) to develop learners' mental representation (i.e., awareness) of these forms (de Graaff & Housen, 2009; Ellis, 2012). This can be done either deductively, i.e., learners are given the rules, or inductively, i.e., learners are encouraged to figure out the rule from examples illustrating the rule (DeKeyser, 1995; Ellis, 2008; Erlam, 2005; Hulstijn, 2005). In explicit instruction, learners are cognizant and thus aware of what they are expected to be learning, or at the very least, that they are expected to learn something about the target language (DeKeyser, 1995; Ellis, 2012; Norris & Ortega, 2000). In contrast, implicit instruction seeks to attract learners' attention to instances of language forms as they appear in communicative input, but does not aim to develop any awareness of the underlying rules describing these forms (de Graaff & Housen, 2009; Ellis, 2012).

Of the the two types of instruction, meta-analytic research has found that explicit instruction yields greater learning gains than implicit instructional procedures (Goo et al., 2015; Norris & Ortega, 2000; Spada & Tomita, 2010). For example, in a seminal article, Norris and Ortega (2000) reported that explicit treatments produced, on average, a much larger effect size ($d = 1.13$) than implicit treatments ($d = 0.54$). Similarly, Spada and Tomita (2010), using the same classification of instruction as Norris and Ortega (2000), also found larger effect sizes for explicit instruction than for implicit instruction, and this advantage was consistent across different language target forms and measures. In a more recent meta-analysis that revisited and updated Norris and Ortega's influential article, Goo et al. (2015), too, found that explicit instruction was more effective than implicit instruction ($g = 1.29$ versus $g = 0.77$, respectively). However, findings of a most recent meta-analysis by Kang et al. (2018) suggest that the difference between explicit and implicit instruction is less pronounced than previously reported ($g = 1.11$ versus $g = 1.38$, respectively), and that implicit instruction is superior than explicit instruction with respect to long lasting effects ($g = 1.76$ versus $g = 0.77$, respectively).

2.1.4 Input-based and output-based form-focused instruction

Form-focused instruction has also been classified according to whether the focus is on input or output (Ellis, 2001, 2012; Nassaji, 2017; Schenck, 2017). Input-based instruction refers to instructional options or treatments that require the use or the processing of input. The assumption behind input-oriented techniques is that learners' attention can be drawn to language forms by means of activities whose purpose is to understand input for meaning. Output-based instruction, on the other hand, refers to instruction that aims to trigger attention to forms by means of elicitation and practice of learners' output (Izumi, 2002; Nassaji, 2017; Schenck, 2017).

2.1.4.1 Input enhancement

One type of input-based instruction particular relevant in this dissertation is *input enhancement*. This type of form-focused instruction (Ranta & Lyster, 2018) seeks to raise learners' attention to form by rendering linguistic forms more salient (Polio, 2007; Sharwood Smith, 1991, 1993, 2013). This concept was introduced by Sharwood Smith (1991, 1993) to replace his previous term *consciousness raising* (Sharwood Smith, 1981), a term which seemed to assume that learners consciously processed all the input they were exposed to. Instead, with the term *input enhancement*, Sharwood Smith put the emphasis on what was done to the input as opposed to the learner. Based on the assumption that enhancement promotes noticing (Schmidt, 1990, 2001, 2010; see also Robinson, 1995; Tomlin & Villa, 1994), Sharwood Smith (1991, 1993) originally defined input enhancement as the process of how certain features in the L2 input become perceptually salient for learners. He argued that perceptual saliency could be externally and internally generated. External saliency can be deliberately created by the teacher (or the researcher) by highlighting targeted forms in the input in such way that it draws learners'attention to such enhanced forms (*bottom-up attention*). Internal saliency, on the other hand, refers to the learners' automatic attention to different aspects of the L2, and it is driven by learners' internal mechanisms such as learning individual differences and personal experiences (*top-down attention*; Kreuz & Caucci, 2007; E. S. Park, 2013).

Revisiting the issue, Sharwood Smith (2013, p. 38) recently complains that most researchers on the area have used the term *enhancement* in the narrowest sense. He argues that the term has mostly been equated with *textual enhancement*, that is, the manipulation of different formatting techniques (e.g., **bolding**, *italicizing*, <u>underline</u>, CAPITALIZING) in an L2 text, with the purpose of directing learners' attention to targeted forms. In an attempt to capture the original meaning of the term *enhancement*, Sharwood Smith defines input enhancement as "*[t]he manipulation of selected (usually linguistic) features of the input deemed important by the language teachers or teaching materials creators with the specific aim of speeding up [L2] development*" (italics in the original). It can be understood from this broad definition that the potential of input enhancement to increase *noticeability* (Schmidt, 1990, 2001, 2010) or *salience* (prominence) of L2 input is not only restricted to a perceptual level (i.e., textual enhancement), as traditionally found in the research literature. In this dissertation, the term *input enhancement* refers to any pedagogical intervention intended to draw learners' attention to L2 form either implicitly or explicitly, within a meaningful context (Spada, 1997, 2011).

Assuming his most recent definition, Sharwood Smith (2013) categorizes input enhancement into three different types: perceptual, conceptual, and affective. Perceptual input enhancement aims to manipulate the learners' perception by rendering a linguistic stimulus auditorily or visually more intense, that is, perceptually noticeable or salient. For Sharwood Smith (see also Ranta & Lyster, 2018, for a similar argument), this type of input enhancement is the first step to assist learners in initiating the processing of L2 linguistic material. He argues that perceptual input enhancement is akin to *getting input through past the front door* repeatedly, hoping that there occurs further internal processing of input into intake, which will eventually lead to stable changes or *restructuring* (S. Carroll, 2001) in the learners' L2 system. Conceptual input enhancement, on the other hand, assumes that learners have noticed or comprehended the meaning of the input as a result of perceptual input enhancement, and thus targets learners' understanding. Sharwood Smith explains that this type of enhancement can have two roles: i) it can facilitate the interpretation of linguistic forms (e.g., word meanings) in the input, and ii) it can have a more analytic role in nature. In the latter case, the enhancement is aimed at helping learners to figure out how the linguistic system works. For example, conceptual enhancement could explain a rule such as subject-verb agreement in the target language. Finally, affective input enhancement refers to the type of enhancement that aims at lowering learners' affective filter, that is, less anxiety and a more positive attitude towards learning the target language in order to process the L2 input more effectively (Krashen, 1985). For instance, the language teacher could affectively enhance a particular L2 linguistic item by attributing a positive value to it.

Of the three kinds of enhancement, the perceptual type in the form of *visual input enhancement* of written materials has been the most widely investigated to date (Polio, 2007; Ranta & Lyster, 2018; Sharwood Smith, 2013). Sharwood Smith (2013) attributes this to the fact that this sort of enhancement is amenable to experimental studies. Although most empirical research on visual input enhancement hitherto has centered on the acquisition of grammar (e.g., Alanen, 1995; Cintrón-Valentín & N. C. Ellis, 2015; Comeaux & McDonald, 2018; Della Putta, 2016; Indrarathne & Kormos, 2017; Issa, Morgan-Short, Villegas, & Raney, 2015; Izumi, 2002; Jahan & Kormos, 2015; Jourdenais, 1998; Jourdenais, Ota, Stauffer, Boyson, & Doughty, 1995; S.-K. Lee, 2007; Leow, Egi, Nuevo, & Tsai, 2003; Loewen & Inceoglu, 2016; Meguro, 2017; E. S. Park & Nassif, 2014; H. Park, Choi, & Lee, 2012; Russell, 2012, 2014, 2016; Tolentino & Tokowicz, 2014; J. White, 2015; Winke, 2013; Wong, 2003; Ziegler et al., 2017), some other areas have been increasingly explored, including vocabulary (e.g., Alsadoon & Heift, 2015; Boers, Demecheleer, et al., 2017; Boers, Warren, Grimshaw, & Siyanova-Chanturia, 2017; S. Choi, 2017; L.-l. Huang & Lin, 2014; Khezrlou, 2018; Khezrlou, Ellis, & Sadeghi, 2017; Y. Kim, 2006; Peters, 2012; Peters, Hulstijn, Sercu, & Lutjeharms, 2009; Ryan, Hamrick, Miller, &

Was, 2018; Sonbul & Schmitt, 2013; Szudarski & Carter, 2016) and pragmatics (e.g., Alcón Soler, 2005; Martínez-Flor & Fukuya, 2005; Nguyen, Pham, & Pham, 2017).

However, the results from studies on visual input enhancement have been mixed (see Benati, 2016; Han, Park, & Combs, 2008; S.-K. Lee, 2007; S.-K. Lee & Huang, 2008; Leow, 2009; Leow & Martin, 2018, for reviews). Some studies have found positive effects on L2 learning (e.g., Cho, 2010; Cintrón-Valentín & N. C. Ellis, 2015; Fang, 2016; Issa et al., 2015; Jourdenais et al., 1995; LaBrozzi, 2016; S.-K. Lee, 2007; N. C. Ellis & Sagarra, 2010; Peters, 2012; Russell, 2012, 2014, 2016; Shook, 1994; Simard, 2009; Szudarski & Carter, 2016; Tolentino & Tokowicz, 2014; Ziegler et al., 2017), whereas others have found no significant or minimal effects (e.g., Alanen, 1995; Della Putta, 2016; Indrarathne & Kormos, 2017; Izumi, 2002; Jahan & Kormos, 2015; Jourdenais, 1998; S. Kubota, 2000; Leow, 1997; Leow et al., 2003; Loewen & Inceoglu, 2016; Meguro, 2017; E. S. Park & Nassif, 2014; Ryan et al., 2018; Winke, 2013; Wong, 2003).

There are also a number of studies that have look at the effect of visual input enhancement on form learning and reading comprehension [2] simultaneously (Boers, Warren, et al., 2017; S. Choi, 2017; Fang, 2016; Jung, 2016; Khezrlou, 2018; Khezrlou et al., 2017; LaBrozzi, 2016; S.-K. Lee, 2007; Leow, 1997, 2001; Leow et al., 2003; Meguro, 2017; Overstreet, 1998; E. S. Park & Nassif, 2014; H. Park et al., 2012; Winke, 2013). Here, too, the results have been inconsistent (S.-K. Lee & Huang, 2008). Some studies have found that visual input enhancement impedes learners' meaning comprehension (e.g., Boers, Warren, et al., 2017; S. Choi, 2017; S.-K. Lee, 2007; Overstreet, 1998; E. S. Park & Nassif, 2014; H. Park et al., 2012); others, however, have reported no such interference effects (e.g., Fang, 2016; Jung, 2016; Khezrlou, 2018; Khezrlou et al., 2017; Leow, 1997, 2001; E. S. Park & Nassif, 2014; H. Park et al., 2012). It has been suggested that the detrimental effect of visual input enhancement on meaning comprehension may

[2]Following SLA literature, the term *form learning* in the present dissertation is used to refer to what learners learn about the target linguistic form (e.g., vocabulary) and the term *comprehension* is employed to refer to what they learn about the informational content or meaning of the input (Leow, Hsieh, & Moreno, 2008; E. S. Park, 2013).

stem from a trade-off between learning the targeted language form and recalling information from the reading materials, i.e., learners' attentional cognitive resources become depleted when attempting to process the enhanced language forms and, as a consequence, the comprehension of meaning is compromised (Boers, Warren, et al., 2017; S. Choi, 2017; S.-K. Lee, 2007; H. Park et al., 2012). One important individual variable that may account for these contradictory results is learner reading proficiency, a relatively neglected factor in the input enhancement literature (Winke, 2013). Given that proficiency level significantly affects L2 comprehension (Carrell, 1991; Fecteau, 1999; K. Kim & Clariana, 2017; J.-w. Lee & Schallert, 1997; Pae, 2018), it is reasonable to assume that high proficient readers would be able to perform better than low proficient readers, regardless of the visual enhancement of the reading materials; more proficient readers may have the sufficient linguistic knowledge and strategies needed to level out any interference from the visually enhanced reading materials (Leow et al., 2008; VanPatten, 1990).

Despite over two decades of research, the effectiveness of visual input enhancement as a pedagogical tool for L2 language learning still remains unclear. It has been suggested that the disparity of results can largely be attributed to methodological variations, such as the amount of exposure to enhanced forms, the number of enhanced items, the length of reading materials, the duration of treatment, the language domain (comprehension vs. production), the measures to assess learners' performance, and the number of participants, among others (Han et al., 2008; S.-K. Lee & Huang, 2008; Meguro, 2017; Ziegler et al., 2017). Overall, meta-analytic work has indicated that the effect of visual input enhancement on L2 learning is small (d = 0.22) and that it also has a small negative effect on meaning comprehension (d = -0.26; S.-K. Lee & Huang, 2008, see also Han et al., 2008).

2.1.4.2 Output enhancement

As noted earlier, enhancement of L2 input can also occur through "pushing" learners to produce output (Benati, 2017; Izumi, 2002; Schenck, 2017; Swain, 1985, 1995; Takashima & Ellis, 1999). Language production can serve as a trigger for internalization of language features, as it requires more attentional resources than comprehension (Boiteau, Malone, Peters, & Almor, 2014), and thus yielding deeper processing of input (Craik & Tulving, 1975; Hopman & MacDonald, 2018; Hulstijn & Laufer, 2001; Schmidt, 2001, 2010). For instance, Izumi (2002) compared textual input enhancement versus output production during reading for comprehension. Output production was operationalized as note taking on passages of the text that participants deemed important during reading. The results indicated that the output production group outperformed the control and the textual enhancement groups. Izumi concluded that language production prompted learners to notice not only the target forms more, but the mismatches between their interlanguage (i.e., developing system) and the target language. Izumi's findings align perfectly with the idea that more noticing to language forms and categories leads to more learning (Schmidt, 2001, 2010), and that output assists learners in realizing the gap between their developing system and the L2 language (Swain, 1985, 1995). Following Schenck (2017), the term *output enhancement* in this dissertation is used to refer to "any form of instructional treatment designed to enhance quality of output" (p. 177).

2.1.5 Summary

This section has focused on the effectiveness of instruction on L2 development. It reported that instruction is not necessary, but beneficial for L2 acquisition. It also described that instruction has generally been classified according to whether the focus is on meaning only (meaning-focused instruction) or meaning and attention to linguistic form

(form-focused instruction). It presented further subcategorizations of form-focused procedures that include implicit and explicit instruction and input-based and output-based instruction. It discussed meta-analytic work comparing implicit and explicit instruction, in which the latter has consistently been found to be more effective. It reviewed research on input-based and output-based instruction, namely input and output enhancement. The review indicated that previous research on the area of visual input enhancement, in particular, has yielded inconsistent or contradictory findings with respect to its effectiveness on L2 learning outcomes.

As will be seen below, the extent to which learners benefit from different instructional conditions may be largely dependent on their individual abilities, such as the cognitive ability to maintain and store information simultaneously (Baddeley, 2017; Ellis, 2012). In this respect, Spada argues that:

> [w]hile there has been extensive research on individual differences and SLA on the one hand, and considerable research on the effects of instruction on SLA on the other, there has been little research on the interaction between individual and instructional variables and their combined effects on learning outcomes. (2011, p. 237)

Moreover, research on this unexplored area that may also help elucidate the effect of visual input enhancement (Biedroń & Pawlak, 2016; DeKeyser, 2012b, 2016; Doughty, 2001, 2003; LaBrozzi, 2016; Meguro, 2017; H. Park et al., 2012; Robinson, 2005; Sawyer & Ranta, 2001; Schenck, 2017; Spada, 1997, 2011; Trofimovich, Lightbown, & Halter, 2013; Vatz et al., 2013).

2.2 Individual differences

Broadly speaking, individual differences are defined as "enduring personal characteristics that are assumed to apply to everybody and on which people differ by degree" (Dörnyei, 2005, p. 4). In psychology, they have been a major subject of research for many years, given their predominant role on many domains of human behavior (see Sackett, Lievens, Van Iddekinge, & Kuncel, 2017, for a recent review). In first language (L1) acquisition, individual differences have been shown to be a critical source of variation in development, and as such, they are regarded as the norm rather than the exception (E. Bates, Dale, & Thal, 1995; Bornstein & Putnick, 2012; Kidd et al., 2018).

In L2 learning, individual differences are also considered to be crucial factors that may account for learners' differential rate of acquisition and ultimate level of attainment (Biedroń & Pawlak, 2016; Dörnyei, 2003, 2005; Ellis, 2012; Gregersen & MacIntyre, 2013; Grey, Williams, & Rebuschat, 2015; Hamrick, 2015; Kormos, 2013; Lado, 2017; Larsen-Freeman, 2018; Li, 2015, 2016, 2017a, 2017b; Robinson, 2001a, 2002; Skehan, 1991, 2012; Tagarelli et al., 2015, 2016). In fact, research has shown that they are the most important predictors of L2 learners' success (Dörnyei, 2005; Kormos, 2013).

Individual difference factors in L2 learning include age (DeKeyser, 2012a; Pfenninger & Singleton, 2017), anxiety (Dewaele, 2017; Dewaele & Al-Saraj, 2015), language aptitude (Lado & Sanz, in press; Skehan, 2012; Wen, Biedroń, & Skehan, 2017), motivation (Csizér, 2017; Dörnyei & Ushioda, 2012), personality (Grey et al., 2015; J. Kim & Nassaji, 2017), and working memory (Juffs & Harrington, 2011; Wen, in press), among others. These individual learner variables have been typically grouped into affective, motivational, and cognitive factors (Ellis, 2012; Kormos, 2013). In this dissertation, the focus is on the latter group, specifically, on working memory and declarative memory.

2.2.1 Types of individual difference research in L2

As noted earlier, there has been abundant research examining the role of individual variables on L2 development (see DeKeyser & Koeth, 2011; Dörnyei, 2005; Gregersen & MacIntyre, 2013; Hamrick et al., 2018; Juffs & Harrington, 2011; Kormos, 2013; Li, 2015, 2016, 2017b; Linck, Osthus, Koeth, & Bunting, 2014; Pawlak, 2017; J. N. Williams, 2012; Wyner & Cohen, 2015, for reviews). This research has been divided into two broad categories according to their nature, namely predictive and interactional studies (Ellis, 2012; Li, 2015, 2017b). Predictive studies are by far the most numerous and they aim to ascertain the relationship between learner individual variables and ultimate learning outcomes, regardless of instructional conditions (Ellis, 2012; Li, 2015, 2017b). In this sense, these studies are correlational and are based on the premise that it is not necessary to adapt instruction in accordance with learners' cognitive profiles (Li, 2017b). These studies examine, for instance, the correlation between general L2 proficiency or any specific language aspect, such as reading, and a given individual difference variable, such as working memory (e.g., Harrington & Sawyer, 1992; Hummel, 2009; Jung, 2018; Sagarra, 2017; Zhou, Rossi, Li, et al., 2017).

Interactional studies, on the other hand, explore the interaction between individual differences and instructional conditions, based on the assumption that the effectiveness of instruction is influenced by individual learner factors (Cronbach & Snow, 1977; Ellis, 2012; Jonassen & Grabowski, 1993; Li, 2015, 2017b; Mackey, 2017; Snow, 1991; Vatz et al., 2013). As such, these studies are experimental, wherein the effects of instructional treatments, such as implicit and explicit instruction (e.g., Carpenter, 2008; Robinson, 1997; Tagarelli et al., 2016) and different types of corrective feedback (e.g., Sheen, 2007, 2010; Yang, Zhang, & Cheng, 2018; Yilmaz & Granena, 2016) are measured and examined in the relation to the individual differences under investigation (Ellis, 2012; Li, 2015, 2017b; Mackey, 2017). This line of individual difference L2 research is inspired by the *aptitude-treatment interaction* paradigm from educational psychology (Burns, Davidson, Zaslofsky, Parker, & Maki, 2017; Corno & Snow, 1986; Cronbach, 1957; Cronbach & Snow, 1977; Fuchs et al., 2014; Lehmann, Goussios, & Seufert, 2016; Lemons et al., 2017; Snow, 1991; Sternberg, Grigorenko, Ferrari, & Clinkenbeard, 1999). In the same vein, this dissertation falls within this category of research.

2.2.2 Aptitude-treatment interaction

According to the aptitude-treatment interaction approach, individual differences moderate (i.e., interact with) the effects of instruction (Cronbach & Snow, 1977; Ellis, 2012; Jonassen & Grabowski, 1993; Kormos, 2013; Larsen-Freeman, 2009; Mackey, 2017; Roehr, 2012; Snow, 1991, 1992; Vatz et al., 2013). Aptitude-treatment interaction effects refer to the notion that some types of instructional techniques (treatment) may be more or less effective for particular learners depending on their individual abilities (aptitude; Cronbach & Snow, 1977; Snow, 1991, 1992).

In this dissertation, *aptitude* refers to relatively stable cognitive abilities[3] that vary from individual to individual, and that may enable a person to perform better at a specific learning area under certain instructional conditions (Buffington & Morgan-Short, in press; Ellis, in press; Wen et al., 2017); *treatment* refers to specific instructional conditions; and *interaction* refers to the differential effects of treatment among individuals that depend upon their particular aptitudes (Snow, 1991; Vatz et al., 2013). The efficacy of a treatment is thus examined across different instructional contexts by looking at the interaction between scores from measures of proficiency and scores from aptitude measures (Li, 2017b; Vatz et al., 2013).

There are at least two ways to investigate aptitude-treatment interaction effects in SLA (Ellis, 2012; Vatz et al., 2013). One way is to employ a classical aptitude-treatment interaction research design, whereby two different instructional conditions or treatments are examined (e.g., inductive vs. deductive instruction). In this case, learners are intentionally matched or mismatched to a treatment condition according to their aptitudes or abilities (e.g., Perrachione, Lee, Ha, & Wong, 2011; Wesche, 1981); those in the matching

[3]These cognitive abilities are *relatively* stable in the sense that they may be less susceptible to change over time due to, for example, training or experience (e.g., working memory, declarative memory). Moreover, the relative stability is given in comparison to more inherently dynamic individual characteristics, such as motivation and anxiety (Biedroń & Pawlak, 2016; DeKeyser & Koeth, 2011; Snow, 1991; Vatz et al., 2013).

condition are hypothesized to outperform those in the mis-matching condition (Cronbach & Snow, 1977; Ellis, 2012; Vatz et al., 2013). However, pure match-mismatch aptitude-treatment interaction design studies are difficult to undertake outside the lab, and thus are less common in SLA (Ellis, 2012; Vatz et al., 2013).

A second, more practical and more frequent way to test aptitude-treatment interaction effects in SLA is to investigate how learners with different cognitive abilities respond to a particular instructional treatment, such as form-focused instruction (Benson & DeKeyser, 2018; Erlam, 2005; Guo & Yang, 2018; Hwu, Pan, & Sun, 2014; Sheen, 2007; Suzuki & DeKeyser, 2017; Yang, Shintani, Li, & Zhang, 2017; Yilmaz, 2013). Here, aptitude measures are not used as grouping, but rather as continuous variables. Studies following this methodology include: i) measures of aptitude, ii) instructional treatment(s), and iii) interactions between learners' abilities and instructional treatment(s), as gauged by language outcome measures (Vatz et al., 2013). The present investigation adopted this second methodological research approach.

Although the number of studies investigating aptitude-treatment interaction relationships in SLA is increasing (see Li, 2015), clearly more research is still warranted (DeKeyser, 2012b, 2016; Ellis, 2012; King & Mackey, 2016; Larsen-Freeman, 2018; Mackey, 2017; Meguro, 2017; Pawlak, 2017; Spada, 2011; Vatz et al., 2013; Wen, 2012, 2016). Echoing other SLA scholars in this regard (e.g., Long, 1983; McLaughlin, 1980; Robinson, 2002, 2007), DeKeyser (2016; see also DeKeyser, 2012b) calls for more research on aptitude-treatment interactions: "[w]hat we need [...] is research that does not ask about the beneficial effect of aptitudes or instructional contexts [...], but that looks at the *interaction* [emphasis added] of two or more of these variables" (p. 358). From a theoretical point of view, this type of interactional research is important because it offers SLA researchers the possibility to draw inferences about the underlying processes in L2 processing and learning (e.g., which aptitudes can facilitate or hamper learning across different instructional contexts). From an instructional, practical perspective, aptitude-treatment interaction research can potentially inform how L2 instruction can be enhanced in order to suit learners' unique needs (DeKeyser, 2012b, 2016; Mackey, 2017; Robinson, 2001a; Vatz et al., 2013; J. N. Williams, 2015).

This study sought to continue this emerging line of research by examining aptitude-treatment interaction effects in a web-based study, with working memory and declarative memory as the focus of the current investigation. These two cognitive abilities are considered primary sources of individual variability in human cognition, with working memory exerting the greatest influence on how people perform high-level cognitive tasks such as language processing and learning (Bunting & Engle, 2015; Kyllonen & Christal, 1990). Moreover, in recent conceptualizations of L2 aptitude as a construct comprised of cognitive abilities (Ellis, in press; Wen et al., 2017), both working memory and declarative memory have been suggested as components of foreign language aptitude (Buffington & Morgan-Short, in press; Robinson, 2007; Skehan, 2016; Wen, in press).

2.2.3 Working memory

Working memory has been defined in several ways (see Cowan, 2017). For example, Baddeley (2017) defines it as "an integrated system involving both temporary storage and attentional control" (p. 301). For Rhodes and Cowan (in press), working memory is the system responsible for "holding a small amount of information to support complex thought" (p. 3). Despite the varying definitions found in the literature, they all coincide in describing working memory as an attentional limited-capacity system that supports complex cognitive processing (Baddeley, 2007, 2017; Baddeley & Hitch, 1974; Cowan, 1988; Daneman & Carpenter, 1980; Hicks, Foster, & Engle, 2016; Linck et al., 2014; Rhodes & Cowan, in press; Unsworth & Engle, 2007).

Several models have also been put forward to explain the functioning of working memory (e.g., see Baddeley, 2007; Baddeley & Hitch, 1974; Cowan, 1988, 2010; Engle, 2002; Truscott, 2017; Unsworth & Engle, 2007). One of the most influential models is the multicomponent cognitive framework proposed by Baddeley and his collaborators (e.g., Baddeley, 2000, 2007; Baddeley & Hitch, 1974). According to this framework, working memory is a system that comprises three components, two of which are "slave" systems that are assumed to be responsible for temporarily holding and processing both verbal (the phonological loop) and spatial (the visualspatial sketchpad) information, respectively. These two domain-specific subsystems are believed to account for the storage functions of working memory (Baddeley, 2017; Li, 2017a; J. N. Williams, 2012). A third, domain-general, limited-capacity component is the central executive, which is regarded as responsible for controlling and regulating attention. This latter component represents the executive functions (e.g., updating information, task-shifting, and inhibition; Miyake & Friedman, 2012; Miyake et al., 2000) of working memory. A relatively recent addition to the model is the episodic buffer component that serves as a link between the slave systems and long-term memory (Baddeley, 2000).

Assuming more contemporary models that emphasize the role of executive functioning (i.e, central executive) as the primary source of individual differences in working memory (e.g., Engle, 2002; Kane, Conway, Hambrick, & Engle, 2007; Linck & Weiss, 2015), and also in line with recent research in SLA (e.g., Faretta-Stutenberg & Morgan-Short, 2018; Suzuki & DeKeyser, 2017; Wen, Borges Mota, & McNeill, 2015), *working memory* in the current investigation is operationalized as an attentional control mechanism that stores and provides access to immediate information necessary for ongoing cognitive processing. *Working memory capacity*, on the other hand, is defined as the ability to simultaneously store and manipulate incoming information (Baddeley, 2007, 2017; Conway, Jarrold, & Kane, 2007; Linck et al., 2014; Miyake & Friedman, 2012; Miyake et al., 2000; Sharwood Smith, 2017; Wilhelm, Hildebrandt, & Oberauer, 2013).

2.2.3.1 Measures of working memory

Working memory has been measured through simple and complex tasks (see Conway et al., 2005; Juffs & Harrington, 2011; Peng et al., 2017, for a survey). Simple tasks (e.g., digit span, letter span) involve recalling short lists of items, and aim to assess the storage components of working memory (H. Bailey, Dunlosky, & Kane, 2011; Juffs & Harrington, 2011). Complex tasks, such as the operation span task (OSpan; Turner & Engle, 1989), on the other hand, entail remembering stimuli (words, digits, letters) while performing other cognitive activities at the same time, and are thought to tax both processing (attention) and storage (memory) components of working memory (Baddeley, 2000; H. Bailey et al., 2011). As such, these latter tasks predict complex behaviors (e.g., reading comprehension, speech production) that require the ability to maintain attention at a maximum state under interference conditions (Conway et al., 2005).

In this dissertation, a complex task, namely the OSpan was used as a measure of working memory capacity for the following reasons. First, complex, language-independent tests such as the OSpan task have been recommended as more accurate measures to examine the association between working memory and L2 processing and learning (Conway et al., 2005; Mitchell, Jarvis, O'Malley, & Konstantinova, 2015; Sanchez et al., 2010; Zhou, Rossi, & Chen, 2017; Zhou, Rossi, Li, et al., 2017), even at advanced levels of proficiency (Yang et al., 2017). Second, in the case of the OSpan task in particular, and in comparison to other common complex tasks such as the reading span, the OSpan task has been found to be a more accurate assessment of working memory capacity in nonnative speakers of English, the target population in this current study (Sanchez et al., 2010; Zhou, Rossi, Li, et al., 2017). Third, the OSpan task has been found to be a valid and reliable measure of working memory capacity (Conway et al., 2005; Unsworth, Heitz, Schrock, & Engle, 2005). And fourth, the OSpan task has been extensively used in SLA (e.g., Faretta-Stutenberg & Morgan-Short, 2018; Goo, 2012; Tolentino & Tokowicz, 2014; Yilmaz, 2013).

It is worth pointing out that in the current study an adapted version of the OSpan task was used, whereby the stimuli to remember were visual (i.e., Klingon characters; Hicks et al., 2016) rather than verbal (i.e., letters), as it is the case in the classic version of the test (see Chapter 4). However, as a complex task, this version of the OSpan equally taxes the executive functions of working memory regardless of the modality of the stimuli, as this type of tests measures domain-generality rather than domain-specificity (Conway et al., 2005; Kane et al., 2004; see also Hamrick et al., 2018, for a discussion on these topics and language acquisition).

2.2.3.2 Working memory and L2 learning

Given its critical role in complex cognitive tasks such as language learning, comprehension and production (Baddeley, 2003, 2017; Cowan, 2015, 2016; Peng et al., 2017), there exists extensive research on the potential effects of working memory on L2 learning (see DeKeyser & Koeth, 2011; Ellis, 2012; Juffs & Harrington, 2011; Li, 2017a, 2017b; Rankin, 2017; Wen, 2012, 2015, 2016; J. N. Williams, 2012, for reviews). These effects have been examined in different areas of SLA, including grammar (e.g., Denhovska, Serratrice, & Payne, 2018; Indrarathne & Kormos, 2018; Tagarelli et al., 2015, 2016), vocabulary, (e.g., Hummel, 2009; Kempe, Brooks, & Christman, 2009; Martin & N. C. Ellis, 2012; Yang et al., 2017), reading comprehension (e.g., Alptekin, Erçetin, & Özemir, 2014; Medina, Callender, Brantmeier, & Schultz, 2017; Sagarra, 2017), and both oral (e.g., Ahmadian, 2012; O'Brien, Segalowitz, Freed, & Collentine, 2007) and written (e.g., Révész, Michel, & Lee, 2017; Zabihi, 2018) production.

Working memory is thought to assist L2 learners with processing simultaneously meaning, form, and use of language forms (e.g., Li, 2017a; J. N. Williams, 2012). More specifically, working memory is involved in key cognitive processes such as attention control, analogical reasoning, explicit deduction, information retrieval and decision making (Li, 2017a; Tagarelli et al., 2016). Moreover, working memory is also crucial for holding metalinguistic information while comprehending and producing L2 language (Roehr, 2008). In this sense, meta-analytic work has shown that working memory has an important role in explaining variability in L2 learning (Grundy & Timmer, 2017; Jeon & Yamashita, 2014; Linck et al., 2014). For instance, in a recent meta-analysis, Linck et al. (2014) reported that, overall, working memory was positively associated with L2 learning ($r = 0.25$). The relationship was somewhat stronger when executive working memory, like in the current study, was examined ($r = 0.27$).

The central role of working memory in L2 learning has been reported in meta-analyses for both comprehension and production (e.g., Jeon & Yamashita, 2014; Linck et al., 2014). Regarding comprehension, working memory has been consistently shown to play a role in influencing outcomes (e.g., Jeon & Yamashita, 2014; Linck et al., 2014). For example, Linck et al. (2014, p. 873) found that working memory has a positive impact on L2 comprehension outcomes ($r = 0.24$). Likewise, Jeon and Yamashita's (2014) meta-analysis also showed that working memory is related to L2 reading comprehension ($r = 0.42$). It is worth noting that this relationship may be moderated by learners' L2 proficiency level. For example, Walter (2004) found that having more working memory capacity helped lower proficiency learners more than higher proficiency learners for reading comprehension, suggesting that the role of working memory may be more significant at lower than at higher levels of proficiency (see Serafini & Sanz, 2016).

Concerning L2 production, meta-analytic research has, too, indicated a significant association with working memory (e.g., Linck et al., 2014). In this case, Linck et al. (2014, p. 873) found an even higher positive correlation for productive than for receptive L2 outcomes ($r = 0.27$).

All together, results from meta-analytic studies suggest that learners with larger working memory capacity are more cognitively adept to meet the demands of learning a foreign language (see also Combs, 2005; Doughty, 2001; Linck & Weiss, 2015; Robinson, 1995). Moreover, the positive relationship between working memory and L2 learning is to be found in both comprehension and production.

Working memory may also account for differential learning gains in L2 instruction that requires learners to alternate attention between meaning and form, such is the case of form-focused instruction through visual input enhancement. In this case, working memory can help learners both notice and process the enhanced forms (Combs, 2005; Doughty, 2001; Indrarathne & Kormos, 2018; Robinson, 1995). Consequently, learners with higher working memory should benefit from visual input enhancement more, one of the instructional conditions in the current dissertation, as they are more capable to deal with attentionally demanding learning tasks such as this one (Linck & Weiss, 2015).

Of primary interest for this dissertation are those studies that have investigated the role of working memory in L2 learning and its interaction with different types of instruction, that is, studies that have looked at aptitude-treatment interactions (e.g., Ando et al., 1992; Brooks, Kempe, & Sionov, 2006; D. M. Chun & Payne, 2004; Denhovska & Serratrice, 2017; Faretta-Stutenberg & Morgan-Short, 2018; Goo, 2012; Indrarathne & Kormos, 2018; Jung, 2017; Kachinske, 2016; Lado, 2017; Li Shuay, 2017; Malone, 2018; Sanz, Lin, Lado, Stafford, & Bowden, 2016; Suzuki & DeKeyser, 2017; Tagarelli et al., 2015, 2016; Yang et al., 2017, 2018; Yilmaz, 2013). For instance, in a study conducted in a classroom-based context, Ando et al. (1992) detected an aptitude-treatment interaction when investigating L2 grammar learning under different instructional conditions. Specifically, the researchers found that learners with high with working memory capacity (as assessed through complex span tasks) benefited more from form-focused instruction (i.e., explicit teaching), whereas learners with low working memory capacity benefited more from meaning-focused instruction (i.e., communicatively-based, implicit teaching).

Tagarelli et al. (2015) also explored the relationship between working memory and L2 learning under two different instructional conditions (incidental and intentional). Participants in both conditions listened to 120 sentences from a semi-artificial language containing English words and German word order, and judged the sentences' semantic plausibility. Participants in the intentional group had to additionally search for rules. Learning was assessed via a grammatical judgment task (GJT) after training. Working memory capacity was measured through the OSpan task and the letter-number serial order task. Results showed that working memory was only predictive of learning on grammatical items in the intentional condition, implying an association of working memory with explicit instructional conditions.

Following a similar methodological paradigm, Denhovska and Serratrice (2017) also examined the potential interaction between working memory and instructional conditions. In this study, however, participants were exposed to written stimulus sentences generated from a natural language, namely Russian, under incidental and intentional instructional conditions. However, unlike Tagarelli et al.'s (2015) study, participants in the intentional group were given metalinguistic information instead of being asked to search for rules. Both receptive and productive grammatical knowledge were measured using a grammaticality judgment task and a fill-in-the-blank task, respectively. Working memory was tested via the OSpan task and the reading span task. The results showed that working memory was related to learning outcomes in the incidental condition, but only in the production domain, suggesting that working memory may play a role in the acquisition of productive knowledge under instructional treatments wherein learner's attention is not directed to linguistic form, such as in meaning-focused instruction.

Another study that investigated how working memory interacts with different types of instruction was undertaken by Indrarathne and Kormos (2018). Intermediate L2 English learners received exposure to a grammatical form (i.e., *causative had*) under implicit and explicit instructional conditions. In the implicit conditions, one experimental group read materials wherein the target form appeared textually enhanced, whereas the other group read materials in which the target form was not textually enhanced; no information about the target form was given to either group. In the explicit conditions, both experimental groups were asked to pay attention to the textually enhanced form, but one group, however, was additionally provided with metalinguistic information. In a pre- and posttest design, comprehension and production of the target form was assessed by means of an auditory GJT and a sentence reconstruction task, respectively. Working memory was examined using four different tasks (the forward Digit-Span test, the Keep Track task, the Plus Minus task, and the Stroop task). To explore the relationship between working memory and instructional conditions, a composite working memory score was used. Results indicated that working memory was associated with both receptive and productive gains in all four instructional conditions. However, the association was much stronger for receptive knowledge, and less so for productive gains in the implicit conditions; the role of working memory was larger in the production domain under the explicit conditions. These findings suggest that working memory can predict both receptive and productive grammar learning, with a stronger link to L2 outcomes in the more explicit instructional conditions. Moreover, according to Indrarathne and Kormos, learners with high working memory may have an advantage with respect to developing receptive knowledge of L2 grammar.

So far, however, most aptitude-treatment interaction interaction studies in SLA have been centered on grammar (see Roehr, 2012; Vatz et al., 2013, for reviews), whereas only a few studies have examined the role of working memory in instructed L2 lexical learning (e.g., D. M. Chun & Payne, 2004; Malone, 2018; Yang et al., 2017, 2018). One of these latter studies was recently carried out by Malone (2018), who investigated the relationship between instructional treatments aiming at promoting incidental L2 vocabulary learning and working memory. Intermediate-level English learners were exposed to low frequent English target words embedded in stories under four different instructional conditions. The experimental conditions differed with respect to the number of exposures (2-4) to the target words, and whether learners received aural input enhancement by listening to the stories while reading or not. In all treatment conditions, participants were asked to read the stories and answer comprehension questions. Knowledge of both word form and word meaning was assessed. Simple (the nonword span task) and complex (the OSpan and the Shapebuilder tasks) working memory tests were used. A composite working memory variable was employed to examine aptitude-treatment interaction effects. Such an effect was found for form recognition. Particularly, working memory was predictive of learning outcomes only in the conditions receiving aural input enhancement.

This results revealed that the effectiveness of (aural) input enhancement as a pedagogical procedure was dependent on learners' working memory abilities.

In a classroom-based study, Yang et al. (2017) investigated whether the effects of different types of L2 vocabulary instruction were mediated by working memory. Specifically, the researchers examined the interaction between working memory and the effectiveness of post-reading word-focused activities as a form of input enhancement. After a reading a text with eight target words glossed in the margin and answering reading comprehension questions, advanced-level English learners were asked to either i) write sentences containing the target words (Sentence writing group), or ii) complete a gapped summary of the text with the target words (Gap-fill group), or iii) answer an essay question without receiving any explicit instruction to use the target words and without being asked to focus on these items (Comprehension-only group). Receptive lexical knowledge was assessed before, immediately after and at a month following the treatments. The reading span task was used to measure working memory capacity. Results showed that working memory mediated the learning gains of the Comprehension-only and Gap-fill groups in the immediate test, suggesting that working memory may be implicated in the acquisition of L2 receptive, not productive lexical knowledge at advanced levels of proficiency under both meaning-focused and form-focused instructional conditions.

In sum, working memory has a central role in accounting for differential learning outcomes in second language acquisition. Overall, this cognitive ability seems to mediate the effectiveness of both meaning-focused (Denhovska & Serratrice, 2017; Indrarathne & Kormos, 2018; Yang et al., 2017) and form-focused (Ando et al., 1992; Indrarathne & Kormos, 2018; Malone, 2018; Tagarelli et al., 2015; Yang et al., 2017) instruction. In regard to explicitness of instructional context, working memory seems to be related to both implicit (meaning-focused) and explicit (form-focused) settings, although the link may be stronger to the more explicit ones (see Baars, 1993; Erçetin & Alptekin, 2013; Linck & Weiss, 2015; Santamaria & Sunderman, 2015; Tagarelli et al., 2016). Of significance for the present dissertation is the fact that there is empirical evidence that the benefits of enhancing input for L2 learning are also moderated by individual differences in working memory capacity (e.g., Indrarathne & Kormos, 2018; Malone, 2018; Yang et al., 2017).

2.2.4 Declarative memory

Another cognitive ability that may also explain distinct effects of L2 instruction is declarative memory. As noted earlier, this construct has been recently suggested as one of the cognitive abilities that may comprise foreign language aptitude (see Buffington & Morgan-Short, in press). In the present study, declarative memory capacity is understood as the ability to consciously recall and use information (Reber, Knowlton, & Squire, 1996; Squire, 2004).

Declarative memory is thought to depend on the declarative memory system, one of the long-term cognitive systems in the brain that is primarily implicated in the processing, storage, and retrieval of information about facts (semantic knowledge) and events (episodic knowledge; Eichenbaum, 2004; Hamrick et al., 2018; Squire, 2004). Learning occurring under this system is posited to be quick, intentional, attention-driven, and predominantly explicit (Buffington & Morgan-Short, in press; M. M. Chun, 2000; Knowlton, Siegel, & Moody, 2017; Morgan-Short & Ullman, 2012; Ullman & Pullman, 2015). Moreover, learning through the declarative memory system is argued to result after a single exposure (e.g., to a word-meaning association), albeit this learning is strengthened by additional exposure (Lum, Conti-Ramsden, Page, & Ullman, 2012; Ullman & Lovelett, 2018).

2.2.4.1 Measures of declarative memory

Declarative memory capacity can be assessed through recall and recognition tasks (Buffington & Morgan-Short, in press; Morgan-Short, Faretta-Stutenberg, Brill-Schuetz, Carpenter, & Wong, 2014; see Hamrick et al., 2018, for a survey). In L2 acquisition research, it has been measured by means of cognitive tasks that can be verbal, such as the paired associates subtest of the Modern Language Aptitude Test (MLAT5; J. B. Carroll & Sapon, 1959) and the LLAMA-B (Meara, 2005); nonverbal, such as the Continuous Visual Memory Task (CVMT; Trahan & Larrabee, 1988); and nonverbal/verbal, such as the visual-auditory learning subtest of the Woodcock-Johnson III Tests of Cognitive Ability (Woodcock, McGrew, Mather, & Schrank, 2001).

In paired associates tasks, participants are asked to memorize pairs of lexical items and then to recall one of the members of the pair when the other is given (Greenberg, Burke, Haque, Kahana, & Zaghloul, 2015). In visual memory recall tasks, participants are required to commit to memory complex figures in a sequence, for example, and then to indicate whether they had seen the figures in the sequence before (Trahan & Larrabee, 1988). And in the nonverbal/verbal, participants are instructed to learn to associate abstract pictures with orally presented words (Ettlinger, Bradlow, & Wong, 2014).

In this dissertation, the MLAT5 and the CVMT were used to assess verbal and nonverbal declarative memory capacity, respectively (see Chapter 4, for details). These tasks have been commonly used in SLA research examining the role of declarative memory in L2 learning (e.g., Carpenter, 2008; Faretta-Stutenberg & Morgan-Short, 2018; Morgan-Short et al., 2015, 2014; see also Buffington & Morgan-Short, in press; Hamrick et al., 2018).

2.2.4.2 Declarative memory and L2 learning

There are two well-known neurocognitive frameworks that have linked declarative memory to L2 acquisition (see Buffington & Morgan-Short, in press; Morgan-Short & Ullman, 2012; Ullman, 2015, for a more detailed discussion of these and other frameworks). These frameworks include Ullman's declarative/procedural model (Ullman, 2001, 2004, 2015, 2016) and Paradis' declarative/procedural model (Paradis, 2004, 2009). Both theories make similar claims with respect to the role of declarative memory in L2 learning. However, they differ in the emphasis they assign to implicit and explicit knowledge, with Paradis' model arguing that declarative memory exclusively contains explicit knowledge, whereas Ullman's model contends that declarative memory can hold both implicit and explicit knowledge. For the purposes of the present study, the latter model is assumed, as it has been the theoretical account most often adopted in L2 research (e.g., Antoniou, Ettlinger, & Wong, 2016; Carpenter, 2008; Ettlinger et al., 2014; Faretta-Stutenberg & Morgan-Short, 2018; Hamrick, 2015; Morgan-Short et al., 2015, 2014).

Declarative memory is assumed to be involved in both L1 and L2 acquisition (Hamrick et al., 2018; Morgan-Short & Ullman, 2012; Ullman, 2001, 2004, 2015, 2016). In the latter, it is thought to underlie the learning, storage and processing of L2 grammar (Ullman, 2015, 2016), at least in the earliest phases of acquisition (Hamrick et al., 2018; Morgan-Short et al., 2014). Regarding vocabulary, the focus of the present dissertation, Ullman's declarative/procedural model posits a role for declarative memory in storing conceptual meanings of words, including their phonological forms. In fact, according to this language model, lexical information (e.g., word meanings) can only be learned in the declarative memory system; L2 learners are dependent on this system to learn vocabulary, independently of their proficiency level (Hamrick et al., 2018; Morgan-Short & Ullman, 2012).

Of particular interest here is the fact that declarative memory seems to be associated with the learning and use of arbitrary knowledge of words at the multi-word level as well (Buffington & Morgan-Short, in press; Hamrick et al., 2018; Morgan-Short & Ullman, 2012; Ullman, 2015, 2016; Ullman & Lovelett, 2018). In this regard, it is argued that complex language forms can be acquired as chunks, especially for adult learners, given the rapidness of learning under this system and the development of declarative memory abilities as later learners of the language (Morgan-Short & Ullman, 2012; Ullman, 2004, 2016). Phrasal verbs, the target form in the current study, are considered complex linguistic forms (Garnier & Schmitt, 2015; Larsen-Freeman, 1991; Larsen-Freeman & Celce-Murcia, 2015; Wray, 2013; see Chapter 3). As such, they are regarded as idiosyncratic (nonderivable) language constructions (Schmitt & Redwood, 2011; Side, 1990). Therefore, it is plausible to assume that individual differences in declarative memory may account for the differential learning of English phrasal verbs (Ullman, 2016; Ullman & Lovelett, 2018).

In a very recent meta-analysis study that assessed the different predictions about language acquisition of Ullman's declarative/procedural model, Hamrick et al. (2018) reported that, as predicted, declarative memory was significantly related to grammar learning at lower, not higher levels of L2 experience (r = 0.46), with a large effect size. Unfortunately, it was not possible for the researchers to report any correlations between L2 vocabulary learning and declarative memory, as they were unable to locate any correlational studies that investigated this relationship. This result prompted Hamrick et al. to call for more research on how declarative memory affects L2 vocabulary learning. As described above, declarative memory should predict L2 lexical acquisition regardless of L2 experience or proficiency level.

Similarly, in a soon-to-be-published review article, Buffington and Morgan-Short reported no studies examining how declarative memory is related to L2 lexical acquisition. Therefore, their conclusions about the role of declarative memory in L2 acquisition are based on evidence from research investigating syntactic development, just as in Hamrick et al.'s (2018) meta-analytic review. Buffington and Morgan-Short concluded that declarative memory has a positive role in L2 learning. Crucially, such a role has been shown to be present under different conditions, be it implicit (Morgan-Short et al., 2014), exposure-based (Antoniou et al., 2016), incidental (Hamrick, 2015) or classroom-based (Faretta-Stutenberg & Morgan-Short, 2018). Buffington and Morgan-Short recommend more research that includes other linguistic features, such as lexico-semantic structures. Particularly, they argue for more studies that investigate "understudied *aptitude/treatment interactions* [emphasis added] such as the role of declarative memory in explicit environments" (Conclusions and Future Directions, para. 1).

It is important to note that in all studies reviewed by Hamrick et al. (2018) and Buffington and Morgan-Short (in press), learning was assessed only receptively. Hence, it is unknown whether declarative memory also affects the development of L2 productive knowledge as well.

A study that explored the role of declarative memory in more explicit instructional conditions was carried out by Carpenter (2008). The explicit condition consisted of providing learners with grammatical rule explanations. The results indicated that declarative memory was implicated in the development of syntactical L2 knowledge, at least during the initial stages of learning. These findings are consistent with the notion that explicit conditions increase learning in declarative memory (Poldrack & Packard, 2003; Ullman, 2016; Ullman & Lovelett, 2018).

To summarize, declarative memory is related to L2 learning. Meta-analytic work has shown its predictive ability to explain variation in L2 attainment (Hamrick et al., 2018). Most importantly, this ability has been demonstrated in aptitude-treatment interactions in different instructional contexts. To this day, however, no research has yet been done on the role of declarative memory in L2 lexical vocabulary acquisition (Buffington & Morgan-Short, in press; Hamrick et al., 2018). Therefore, the predictions below (RQ1 and RQ2) are based on research on child language acquisition and theories on declarative memory as a general-purpose learning system.

2.2.5 Summary

This section has concentrated on how individual differences affect second language attainment success. Individual difference research has been classified according to two methodological paradigms: predictive and interactional. The present dissertation follows the latter, wherein the role of cognitive abilities on L2 learning is explored with respect to instructional conditions, that is, aptitude-treatment interactions. Two key cognitive individual difference factors are investigated, specifically working memory and declarative memory. In the case of the former, the review of the literature revealed that it greatly influences L2 learning outcomes, and such an influence has been attested in different areas of SLA, in both receptive and productive domains, and in different instructional situations, including those where L2 input has been enhanced to promote learning. For the latter, the present review indicated that it is positively associated with L2 learning gains, most notably for receptive syntactic abilities.

A clear gap found in the literature of both working memory and declarative memory is the paucity of research examining aptitude-treatment interactions. To fill this research gap, the present dissertation looked at aptitude-treatment effects in a web-based, intelligent computer-assisted language learning experiment targeting English phrasal verbs.

2.3 Intelligent Computer-Assisted Language Learning

Intelligent computer-assisted language learning (ICALL) refers to the integration of computational linguistics into computer-assisted language learning (CALL; Meurers & Dickinson, 2017, see also Heift & Schulze, 2007; Jurafsky & Martin, 2009; Nerbonne, 2003, for a thorough discussion and an overview). Typically, this integration has been achieved through the application of natural language processing (NLP), one of the branches of artificial intelligence (Antoniadis & Desmet, 2016; Heift, 2013; Lu, 2018; Meurers, 2013; Meurers & Dickinson, 2017; Nerbonne, 2003; Nyns, 1989; Schulze &

Heift, 2012). Other branches of artificial intelligence employed in ICALL include learner modeling (i.e., capturing and adapting to the learner's individual characteristics) and expert systems (i.e., using relevant knowledge from [learning] domain experts; see Gamper & Knapp, 2002; Heift, 2013; Heift & Vyatkina, 2017; Levy, 2009; Schulze & Heift, 2012, for review). In the current dissertation, the term *ICALL* is used to refer to NLP-enhanced ICALL.

NLP supports the automatic analysis of human language by applying a variety of techniques. These techniques include tokenizing (i.e., splitting text into meaningful units), part-of-speech annotation (i.e., assigning basic parts of speech labels to words, such as verbs, nouns, adjectives, and so forth), and syntactic parsing (i.e., assigning structure to a sentence), among others (Lu, 2018; Meurers, 2013; K. Petersen & Sachs, 2015).

2.3.1 Uses of NLP in ICALL

As the applied branch of computational linguistics (Meurers, 2013), NLP can potentially have many different uses in ICALL (see K. Petersen & Sachs, 2015, for examples of such uses). More generally, however, NLP uses in the context of language learning can be grouped into two basic types (Meurers, 2013; see also Antoniadis & Desmet, 2016; Lu, 2018). One type concerns using NLP resources for the analysis of learner language, i.e., learners' written or spoken production in the target language. This kind of analysis can be used to identify learners' first language, to assess learners' performance, to annotate learner corpora, or to provide contextual corrective feedback (Ai, 2017; Heift, 2017; Heift & Vyatkina, 2017; Lu, 2018; Meurers, 2013, 2015; Meurers & Dickinson, 2017; Quixal & Meurers, 2016). Instances of this type of NLP use are intelligent tutoring systems. As educational tools (Nyns, 1989), these NLP-based ICALL systems aim at identifying and providing learners with individualized feedback on errors, all within the context of electronic textbooks (Heift & Vyatkina, 2017; Lu, 2018; Meurers, 2013;

Meurers & Dickinson, 2017; K. Petersen & Sachs, 2015). Examples of intelligent tutoring systems are TAGARELA for L2 Portuguese (Amaral & Meurers, 2011), Robo-Sensei for L2 Japanese (Nagata, 2009), E-Tutor for L2 German (Heift, 2010, 2016), and, more recently, FeedBook for L2 English (Rudzewitz, Ziai, De Kuthy, & Meurers, 2017).

A second broad type of NLP use in foreign language learning deals with the analysis of authentic spoken and written texts, i.e., texts aimed at native speakers, not language learners. The analysis occurs at different levels of the linguistic system (e.g., morphology, syntax, and lexicon). The resulting information and annotations can then be used for the automatic selection, sequencing, enhancement of texts as target language input, which, as such, can be rendered salient for learners. The results of this NLP-driven analysis can also be employed for the automatic generation of exercises, such as those embedded in L2 reading materials and language tests, such as multiple-choice cloze tests (Heift & Vyatkina, 2017; C.-L. Liu, Wang, Gao, & Huang, 2005; Lu, 2018; Meurers, 2013; Meurers & Dickinson, 2017; Meurers et al., 2010). ICALL (also called ATICALL; Meurers, 2013) applications that illustrate the second type of NLP use in L2 language learning include WERTi (Meurers et al., 2010), SMILLE (Zilio, Wilkens, & Fairon, 2017), SmartReader (Azab et al., 2013), and FLAIR (Chinkina & Meurers, 2016); the target language in these ATICALL applications is English. In this dissertation, a customized version of the web-based WERTi system (Meurers et al., 2010) was used to investigate the acquisition of English phrasal verbs (see Chapter 5, for details).

2.3.2 Potential of ICALL in language learning and SLA research

Overall, ICALL has a great potential to contribute to language learning and SLA research. In this sense, Meurers (2013) posits that: "the use of NLP in the context of learning language offers rich opportunities, both in terms of developing applications in support of language teaching and learning and in terms of supporting SLA research"

(p. 9). From a theoretical perspective, ICALL offers the opportunity to gain insights into processes and outcomes of SLA through the collection of data that might be otherwise not possible in research conducted in classroom- and lab-based settings. ICALL applications allow for the collection of detailed individual performance logs, which can contribute to deepening the understanding of how incremental changes in L2 development might occur during experimental manipulation. This, in turn, can provide a more fine-grained and nuanced understanding of L2 learning (Meurers, 2013; Meurers & Dickinson, 2017; Ziegler et al., 2017).

From a methodological view, ICALL also offers unique affordances for SLA research. First, web-based ICALL systems like WERTi afford the possibility of investigating instructed SLA in real-life language instructional contexts (e.g., Internet-based instruction), an underesearched area in SLA (MacWhinney, 2017). Second, web-based ICALL systems can be used to address the critical need to deal with the limited amount of exposure to target language features as well as the small amount of target items in input enhancement research, for example (Han et al., 2008; Ziegler et al., 2017). Third, ICALL systems not only can analyze and enhance learning materials on the fly, but can also generate exercises that target both receptive and productive practice (e.g., Meurers et al., 2010). In this way, ICALL is facilitative of instruction of relevant language features in a dynamic fashion (Meurers, 2013; Ziegler et al., 2017). Fourth, ICALL applications can support individual selection of authentic materials by the learner, rendering learner-centered, individualized instruction feasible (Meurers, 2013; Ziegler et al., 2017). Consequently, learners become autonomous and possibly more motivated (Csizér, 2017; Monfared, Cervantes, Lee, & Jackson, 2018), which, in turn, may also make them process, comprehend, and retain L2 input more effectively (Hulstijn, 2001; Loschky, 1994). Fifth, by using web-based ICALL applications to run their experiments, SLA researchers can also address the need of increasing sample sizes and thus gaining statistical power (Plonsky, 2017). As such, web-based ICALL system can significantly reduce the laborious and time-consuming process of collecting data in the lab (Meurers & Dickinson, 2017; Ziegler et al., 2017). Sixth, ICALL systems can facilitate the collection of authentic learner data (i.e., learner corpora; Meurers & Dickinson, 2017). Finally, web-based ICALL systems can provide the framework for an extensive and systematic implementation of testing and treatments, including the testing of individual differences (Meurers, 2013; Meurers & Dickinson, 2017; Ziegler et al., 2017). In sum, ICALL is a valuable tool for language learning and empirical SLA research.

2.3.3 Empirical ICALL research

To date, there has been a dearth of empirical research investigating the integration of NLP-supported ICALL and instructed SLA (e.g., Arispe, 2012; I.-C. Choi, 2016; Heift, 2005; Heift & Rimrott, 2012; Nagata, 1996; K. A. Petersen, 2010; Wilske, 2015; Ziegler et al., 2017). Most of this research has focused on L2 grammar, with the aim to examine the efficacy of providing corrective feedback on learning outcomes and/or learner-computer interaction (e.g., I.-C. Choi, 2016; K. A. Petersen, 2010; Wilske, 2015). On the whole, this research has found that explicit feedback in the form of metalinguistic information is more beneficial for learners. For instance, Wilske (2015) developed an web-based ICALL system that offered corrective feedback on L2 German grammar. Results indicated that metalinguistic feedback was more effective in producing immediate learning gains than implicit feedback via recasts (i.e., reformulation of learners' utterances).

Of relevance here was an ICALL pilot study conducted by Ziegler et al. (2017). The study explored the potential benefits of using ICALL systems to investigate the integration of ICALL and instructed SLA. More specifically, the study examined whether automatic input enhancement supported by NLP-driven analysis of news texts from Reuters could afford the acquisition of implicit and explicit knowledge of English articles. The researchers used a tailored version of the WERTi system developed by Meurers et al. (2010). The WERTi system integrates NLP technologies to support the automatic identification of target forms in webpages. In turn, these identified target forms can automatically become color highlighted (Color), clickable (Click), transformed into fill-in-the-blanks (FIB), or multiple-choice options (MC). The system also allows logging of relevant text properties and all interactions of the learner with the system, making it possible to incrementally track development of abilities. In Ziegler et al.'s study, learners were asked to read visually enhanced webpages containing news content according to four treatment conditions (Color, Click, MC, FIB). The results indicated that the MC group experienced significant gains in explicit knowledge in a pretest-posttest comparison, suggesting that automatic input enhancement has the potential to support L2 learning. Interestingly, the ICALL system also allowed the collection of learner logs during treatment. The results of the analysis of these computer logs indicated that learners' development was both nonlinear and incremental. These findings were obtained due to the distinct methodological affordances offered by ICALL systems.

2.3.4 Summary

In this section, the concept of ICALL, especially NLP-based ICALL, was described. Then, the two broad uses of NLP in ICALL were presented. Following this description, the theoretical and methodological affordances of ICALL for L2 language learning and SLA research were discussed. Lastly, the limited empirical research on the area of ICALL, with emphasis on ATICALL systems was reviewed. In particular, the study conducted by Ziegler et al. (2017) is of importance, as this previous study served as the conceptual framework and starting point for the current dissertation, which aims to contribute to the emerging line of interdisciplinary research at the intersection of CALL, computational linguistics and instructed SLA.

2.4 The present dissertation

The aims of the present dissertation were: (i) to investigate how individual differences in working memory and declarative memory impact the acquisition of L2 lexical knowledge, (ii) to determine whether the potential impact of working memory and declarative memory is modulated by instruction type (meaning-focused versus form-focused), and (iii) to ascertain whether an ICALL system can be used to address these topics. The linguistic focus was on English phrasal verbs.

2.4.1 Research questions

The following research questions informed the present investigation:

RQ1: Do working memory and declarative memory impact the acquisition of English phrasal verbs?

RQ2: Is the potential impact of these individual differences modulated by instruction type (meaning-focused versus form-focused)?

RQ3: Can an ICALL environment provide the affordances necessary to conduct large-scale, experimental-style research addressing these research questions?

Regarding RQ1, based on previous research (e.g., Jeon & Yamashita, 2014; Kempe et al., 2009; Linck et al., 2014; Martin & N. C. Ellis, 2012; J. N. Williams, 2012; Yang et al., 2017), it was predicted that working memory would have an impact on the acquisition of L2 lexical knowledge. In the case of declarative memory, theoretical predictions suggest that declarative memory would influence L2 vocabulary learning. However, as there is no previous research investigating this very topic, a strong prediction could not be made.

For RQ2, it was hypothesized that the type of instruction (meaning-focused versus form-focused) would modulate the possible influence of working memory (Ando et al., 1992; Denhovska & Serratrice, 2017; Indrarathne & Kormos, 2018; Malone, 2018; Tagarelli et al., 2015; Yang et al., 2017), and that the effect would be more likely to occur in the form-focused condition (Ando et al., 1992; Erlam, 2005; Indrarathne & Kormos, 2018; Li, 2017a; Malone, 2018; Tagarelli et al., 2015; Yang et al., 2017). Moreover, learners with higher working memory capacity were expected to better handle the cognitive demands of form-focused instruction (i.e., paying attention to meaning and form simultaneously), and thus to benefit more from it (Linck et al., 2014). 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 considered to be associated with more explicit learning processes (Poldrack & Packard, 2003; Ullman, 2016; Ullman & Lovelett, 2018).

As for RQ3, it was expected that an ICALL system would provide the capabilities necessary to carry out a large scale web-based experiment that would permit address the research questions of this dissertation (Ziegler et al., 2017).

Chapter 3

Selection of linguistic target

This chapter deals with the selection of the linguistic target and contains two parts. The first part reviews relevant literature on the acquisition of phrasal verbs. It also discusses corpus-based research on native and nonnative phrasal verb use. The second part reports the results of a corpus-oriented study supported by NLP technology. This study allowed the selection of learning targets and materials.

3.1 Part I: Phrasal verbs as linguistic target

The linguistic focus in this dissertation is on English phrasal verbs. Phrasal verbs, such as *back up* (support), *go on* (continue), and *put off* (postpone), are very typical and important features of the English language (Armstrong, 2004; Thim, 2012), thereby crucial for L2 learners to master (Garnier & Schmitt, 2016; Ke, 2016). In fact, knowledge of phrasal verbs is taken as indicative of L2 proficiency (Bywater, 1969; Cornell, 1985; Howarth, 1998; D. Liu, 2012; Wyss, 2003). Corpus-based research has shown that these multi-word constructions are highly frequent in language use (Biber, Johansson, Leech, Conrad, & Finegan, 1999; Davies, 2009; Gardner & Davies, 2007; D. Liu, 2011). Biber et al. (1999), for instance, found that phrasal verbs occur almost 2,000 times per million words. Gardner and Davies (2007, p. 347) reported that phrasal verbs occur approximately every 192 words on average. Notwithstanding their high frequency of occurrence, phrasal verbs are notoriously difficult for English learners, even at advanced levels of proficiency (M. Chen, 2013a; Coe, 2001; Cornell, 1985; Garnier & Schmitt, 2015, 2016; Larsen-Freeman & Celce-Murcia, 2015; Neagu, 2007; Trebits, 2009; B. J. White, 2012; Yasuda, 2010). The significance and challenging nature of phrasal verbs render them particularly interesting and worthy language forms to study within L2 research.

3.1.1 Delimitation and definition of phrasal verbs

Apart from the term *phrasal verb*, a look at the literature reveals that the same type of combination has received different names, including *particle verbs*, (e.g., Dehé, 2002), *verb particle combinations* (e.g., Fraser, 1974), and *verb particle constructions* (e.g., Lipka, 1972). According to McArthur (1989, p. 38), the term *phrasal verbs* seems to be "the winning term". *Phrasal verbs* is the term that will be used in this dissertation since it is also the most commonly used label in L2 teaching and learning literature (see, e.g., Larsen-Freeman & Celce-Murcia, 2015; Schmitt & Redwood, 2011).

Phrasal verbs are thought to be at the interface between syntax and semantics (Davies, 2009; Gardner & Davies, 2007; Trebits, 2009). Hence, definitions of phrasal verbs in the literature often combine syntactic and semantic criteria. From a syntactic perspective, phrasal verbs are seen as multi-word verbs that are comprised of a verb accompanied by either a preposition, an adverb or both (e.g., *put up with*; Meurers et al., 2010). From a semantic perspective, phrasal verbs are considered to be a lexical category whose constituents produce different combinations of meanings (Downing & Locke, 2006; Larsen-Freeman & Celce-Murcia, 2015). These meanings can in turn be *literal* (e.g., *sit down* [sit]), *figurative* (e.g., *cut off* [interrupt]), and/or *idiomatic* (e.g., *put off* [tolerate]). Considering both perspectives, one can define phrasal verbs as multi-word expressions which function semantically and syntactically as a single unit. As such, a phrasal verb consists of a verb and one or two elements called *particles* (e.g., *up, down, out, off*; Biber et al., 1999; Carter & McCarthy, 2006; Quirk, Greenbaum, Leech, & Svartvik, 1985). Particles are said to have an adverbial status, meaning that they alter the meaning of the verb to a lesser or greater extend (Ke, 2016; Lipka, 1972; Side, 1990).

Given that the phrasal verbs targeted in this dissertation were automatically detected and extracted from a corpus using NLP methods (see Part II for details), a syntactic rather than a semantic criterion was employed to determine what counted as a phrasal verb.[4] More specifically, the syntactic dependency parser from spaCy (`https://spacy.io/`), an open-source software library, was used. In general terms, a parser's task is to allocate or assign a structure to a given sentence, i.e., *parse* a sentence (Dickinson, Brew, & Meurers, 2013, p. 54). In particular, a dependency parser takes a sentence as a sequence of words, and produces a connected tree with dependency links or arcs between all the words. These arcs represent syntactic relationships between word categories, as shown in Figure 1.

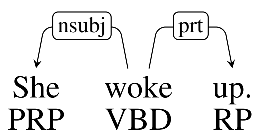


Figure 3.1. Dependency-annotated sentence containing the phrasal verb *wake up*.

spaCy is a statistically-based system, meaning that the part of speech (POS) tags that assigns to the different elements of a sentence is the best estimate based on the language data it was trained on. These tags come from the Treebank Project tagset (Taylor, Marcus, & Santorini, 2003), which is a large annotated corpus of modern English whose notations are commonly used as reference in computational linguistics. Accordingly, the output of spaCy's syntactic processing of a phrasal verb seen in Figure 1 has the *prt* notation to denote the relation (i.e., dependency) between the verb in past tense (*VBD*), and the particle (*RP*). For our purposes, every time the parser's analysis of a sentence yielded the same output as depicted in Figure 1, it was then considered as an instance of a phrasal verb.[5] Of course, these computational tools are not perfect and therefore some tagging inaccuracies can occur, but, on the whole, the accuracy of modern English POS taggers is about 97%, which is about the same as the average human (Manning, 2011). Particularly in the case of spaCy, it has been reported that its dependency parser is one of the most accurate parsers

[4]Gardner and Davies (2007), D. Liu (2011), and Trebits (2009) also adopted a syntactic criterion in their computer-assisted analysis of English phrasal verbs.

[5]The results of the analysis were manually corroborated with the purpose of identifying false positives.

currently available (Honnibal & Johnson, 2015).

3.1.2 Difficulties of learning phrasal verbs

The phrasal verb, which is believed to be the most productive source of new words in English (Bolinger, 1971), is a challenging language form to acquire. In the context of L1 acquisition, English speaking children have difficulties when acquiring phrasal verbs (Diessel & Tomasello, 2005; Gries, 2011; Villavicencio et al., 2012). In L2 acquisition, the difficulties are, of course, much more pronounced. This difficulty is such that learners tend to avoid using phrasal verbs, including those whose L1s have similar counterparts, such as Dutch or Swedish (Dagut & Laufer, 1985; Hulstijn & Marchena, 1989; Laufer & Eliasson, 1993; Liao & Fukuya, 2004).

Phrasal verbs are also regarded as part of *formulaic language*, which refers to sequences of words whose combination produces “a joint grammatical, semantic, pragmatic, or textual effect greater than the sum of the parts” (Wray, 2006, p. 591). As such, the problems of learning phrasal verbs can then be traced back to the fact that they are semantically, syntactically and pragmatically complex (Garnier & Schmitt, 2015; Larsen-Freeman, 1991; Larsen-Freeman & Celce-Murcia, 2015; Wray, 2013). As will be seen below, these problems elicit learners’ incorrect use of phrasal verbs (Chang, 2001; Coe, 2001; J.-A. Lee, 2001; Papaefthymiou-Lytra, 2001; Strong, 2013; Wilson & Wilson, 2001).

3.1.2.1 Semantic difficulties

Semantically speaking, phrasal verbs are challenging for L2 learners to decipher and understand because their meanings are opaque: the meaning resulting from the combination of a verb and a particle is often different from that of the verb occurring alone, as in (1) compared to (2).

(1) She *turned* (moved) her face towards the wall.

(2) She *turned down* (rejected) the job offer.

Learners tend to struggle with the distinction between these two meanings, that is, phrasal versus nonphrasal uses (Meurers et al., 2010), which can produce the omission of the particle (Wilson & Wilson, 2001), as in (3) compared to (4).

(3) *He *put* (wear) the jacket.[6]

(4) He *put* ***on*** the jacket.

Moreover, and unlike native speakers who process phrasal verbs as a single lexical chunk (Cappelle, Shtyrov, & Pulvermüller, 2010; Schunack, 2017), nonnatives find it problematic to recognize that two or more orthographic words, in this case a verb and a particle, function together as a single syntactic and semantic unit (Siyanova & Schmitt, 2007; but see Paulmann, Ghareeb-Ali, & Felser, 2015, who recently presented neurophysiological evidence that proficient L2 learners exhibit similar phrasal verb processing patterns as native speakers do). Instead, they often perceive the association between a verb and a particle to be completely arbitrary (Side, 1990), with the particle in particular seen as illogical (Thibeau, 1999).

Another very important semantic challenge is the fact a large number of phrasal verbs are *polysemous*, that is, they have several possible related meanings or senses (Consigny, 2006; Gardner & Davies, 2007; Garnier & Schmitt, 2016; Mahpeykar & Tyler, 2015). Indeed, according to Gardner and Davies (2007, p. 353), phrasal verbs have approximately 5.6 different meanings on average. Some phrasal verbs, for example, have a transparent, compositional meaning, that is, both the verb and the particle maintain their *prototypical*, literal meanings (Morgan, 1997), as in (5).

(5) He *pulled down* (removed) the curtains.

And a semantic link between different senses can be established (Schmitt & Redwood, 2011), as in (6) and (7).

[6]In linguistics, an asterisk (*) is used to denote an ungrammatical sentence.

(6) She *filled in* (completed) the form.

(7) They *filled in* the detective (told) on what happened last night.

For some other phrasal verbs, however, the connection is weak and their meanings are difficult to interpret, as in (8) and (9).

(8) The guest *put up* (stayed) for a night.

(9) The police *put up* (displayed) a notice.

Undoubtedly, being unable to predict of the meaning of a phrasal verb from the combined literal meaning of its parts is a major source of learning difficulty (Torres-Martínez, 2017; B. J. White, 2012).

3.1.2.2 Syntactic difficulties

From a syntactic point of view, particle placement in phrasal verbs is also problematic for learners. The fact that phrasal verbs can be transitive and intransitive allows for particles to be separated from their verbs by pronouns, adverbs and noun phrases, as in (10), (11) and (12), respectively.

(10) She *pick* them *up* (collected) from the airport.

(11) He *came* straight *over* (visited) to see his family.

(12) I tried to *calm* the younger children *down* (pacify).

Additionally, learners not only need to decide whether the verb is separable, but also what it is possible to separate (e.g., adverb, pronoun, short noun phrase, long noun phrase; Magnusson & Graham, 2011). Not knowing what to separate can, for example, lead to ordering problems like erroneously placing the object pronoun after the particle instead of before (Kadia, 1998; Papaefthymiou-Lytra, 2001), as in (13) compared to (14).

(13) *She *called* ***up him*** (phoned) yesterday.

(14) She *called* ***him up*** yesterday.

For Schmitt and Redwood (2011, p. 174), the decision about verb separability goes beyond grammatical considerations such as transitivity. Rather, the decision often depends on a mixture of "stylistics and syntactic conventions, context, prosody and intended meaning."

3.1.2.3 Pragmatic difficulties

From a pragmatic perspective, phrasal verbs are also troublesome for L2 learners. To begin with, particle movement can be further determined by *information structure*. In this sense, Larsen-Freeman and Celce-Murcia (2015), based on the *principle of dominance* by Erteschik-Shir (1979), posit that if the object noun phrase (NP) is *dominant*, that is, the NP is elaborate, long and represents *new* information, it is likely to occur after the particle as in (14). But if the direct object is short (e.g., a pronoun) and contains *old* information, this is expected to occur before the particle, as in (15). In short, if the object is not dominant, it will occur between the verb and the particle in separable verbs.

(14) She *gave up* **smoking** a long time ago. (New information)

(15) She *gave* **it** *up* a long time ago. (Old information)

Another pragmatic aspect that L2 learners have to deal with is the use of phrasal verbs. Are they only used in informal contexts? Can they be used in academic writing? Corpus-based research show that native speakers use phrasal verbs much more commonly in spoken or informal registers than in written or formal ones (Biber et al., 1999; D. Liu, 2012). Nonetheless, it has been found that phrasal verbs are also present in formal contexts (D. Liu, 2012). Recently, Larsen-Freeman and Celce-Murcia (2015) offer two explanations for the use or nonuse of phrasal verbs in certain contexts. The first pertains to the fact that certain phrasal verbs such as *check out* are so tightly associated with a particular field that there are no comprehensive synonyms. The second concerns with the fact that there are contexts in which the use of Latinate or one-word verbs is preferable to the use of phrasal verbs due to the importance of the information to be conveyed. Examples of this are airport announcements wherein comprehending messages is crucial for international travelers. It is, however, challenging for nonnatives to know when it is more appropriate to use a phrasal verb than a one-word synonym, and vice versa. In general, learners prefer to use synonymous one-word verbs over phrasal verbs (Dagut & Laufer, 1985; Hulstijn & Marchena, 1989; Liao & Fukuya, 2004; Siyanova & Schmitt, 2007; Sung, 2017), making them appear rather bookish and formal.

In sum, learners of English are confronted with many syntactic, semantic and pragmatic nuances that render phrasal verbs extremely difficult to acquire. Of the three aspects described above, it is reasonable to assume that the semantic one is perhaps the hardest for learners to contend with (Hulstijn & Marchena, 1989; Larsen-Freeman, 1991; Neagu, 2007). Therefore, the focus of instruction in this dissertation was on the development of both receptive and productive semantic knowledge of target phrasal verbs through reading short news texts online.

3.1.3 Empirical research on L2 acquisition of phrasal verbs

Despite the importance and difficulty of phrasal verbs, there is little empirical research examining the L2 acquisition of phrasal verbs (Birjandi, Alavi, & Karimi, 2015; Chiang, 2012; Garnier, 2016; Khatib & Ghannadi, 2011; D. Liu, 2003; Magnusson & Graham, 2011; Nassaji & Tian, 2010). The paucity of research can be largely attributed to the complexities and peculiarities of these multi-word constructions, as explicated above. Nonetheless, the number of investigations on the topic fall into two main strands. On the one hand, there are studies that follow a semantic, cognitive linguistic approach (see Lakoff & Johnson, 1980; Randal, 2012). These studies investigate whether encouraging learners to use orientational metaphors (i.e., indicating movements in space), such as COMPLETION IS UP (*eat up, give up*), can be effective on the teaching and learning of phrasal verbs (e.g., Boers, 2000; Condon, 2008; Farsani, Moinzadeh, & Tavakoli, 2012; Ganji, 2011; Kartal & Uner, 2017; Kövecses & Szabcó, 1996; H. Lee, 2016; Neagu, 2007; B. J. White, 2012; Yasuda, 2010).

On the other hand, there exist instructed SLA studies investigating phrasal verb acquisition in relation to the effectiveness of a) instructional conditions (e.g., Chévez-Herra, 2013; Garnier, 2016; Gharedaghi & Touran, 2016; Hare, 2010; Khatib & Ghannadi, 2011; Magnusson & Graham, 2011; Resketi & Bagheri, 2014; Shin & Christianson, 2012); b) input enhancement (e.g., Behzadian, 2016; Birjandi et al., 2015; Chiang, 2012, 2013; Khatib & Ghannadi, 2011; Pam & Karimi, 2016); c) concordance-based or data-driven learning instruction (DDL) instruction (e.g., Azzaro, 2012; Behzadian, 2016; Johns, 1990; E. N. Mohammadi & Tashakori, 2015); d) processing instruction (e.g., Chiang, 2012; Thibeau, 1999); e) type of task (e.g., Khodareza & Shabani, 2016; Nassaji & Tian, 2010); f) provision of positive and negative evidence (e.g., M. Kubota, 1997); g) use of captions (e.g., Pasban, Forghani, & Nouri, 2015); and h) computer-mediated-communication (CMC) instruction (e.g., E. Mohammadi & Mirdehghan, 2014). Taking into account that the present dissertation deals with the acquisition of phrasal verb from the perspective of instructed SLA, the discussion that follows is confined to empirical studies on this area.

As indicated above, some empirical instructed SLA-based studies have looked at whether instructional conditions make a difference on phrasal verb acquisition. For instance, Shin and Christianson (2012) compared the extent to which implicit versus explicit instruction could lead to the development of productive knowledge of syntactically simple phrasal verbs (e.g., *The girl is turning the heater down*). To that end, researchers used the phenomenon of *structural (or syntactic) priming*, which refers to speakers' tendency to repeat the same structural pattern they had previous encountered (Bock, 1986). In a lab setting environment, forty-eight highly proficient Korean learners of English were exposed to phrasal verb constructions under three priming conditions: long-lag, no-lag, and explicitly reinforced (i.e., no-lag plus explicit instruction). *Lag* refers to the number of intervening elements between the prime sentence and the target sentence. Explicit instruction was implemented through rule presentation (DeKeyser, 1995). Participants' productive and receptive knowledge was measured via a picture description task and a grammaticality judgment test, respectively. Results revealed that, regardless of the instructional condition, there was an overall increase in the production of phrasal verb constructions in both immediate and delayed posttests in comparison to a baseline established in the pretest. Crucially, these findings suggest that it is possible to acquire L2 phrasal verb knowledge in an incidental fashion as a by-product of exposure. The latter is what Hare (2010) precisely set out to investigate. Specifically, Hare studied whether phrasal verbs can be acquired incidentally (i.e., without intention) through reading. In a pretest–posttest experiment, 14 intermediate and advanced English as Second Language

(ESL) learners read eight one-page experimentally modified stories with follow-up activities over a three-week period. In total, participants were exposed eight times to the same eight low-frequent, opaque phrasal verbs embedded in the stories, which were chosen by the researcher on the basis of her "native intuition and experience" (p. 45). At pre-test, only receptive knowledge of target items was measured. After the treatment, both receptive and productive knowledge were assessed in immediate and delayed post-tests. Results show that the instructional treatment resulted in learning gains in both types of knowledge, which indicates that phrasal verbs can be learned under incidental conditions.

In the same vein, Resketi and Bagheri (2014) examined incidental acquisition of phrasal verbs by comparing enhanced versus unenhanced extensive reading. Extensive reading (ER) is an approach to L2 reading instruction whereby learners are encouraged to read large amount of texts in the target language (Bamford & Day, 2004). *Enhanced ER* was operationalized as extensive reading plus post-reading activities (e.g., oral book reports, book summaries), whereas *unenhanced ER* referred to reading extensively only. Twenty-five intermediate-level English as a Foreign Language (EFL) students were assigned to either the experimental (enhanced ER) group, or the control (unenhanced ER) group. Both groups took the same 50-item test (multiple-choice and fill-in-the-blank exercises) before and after a five-week instructional treatment. The researchers found that enhanced ER was significantly more effective than unenhanced ER in promoting learners' incidental acquisition of phrasal verbs. Importantly, these results indicate that acquiring phrasal verbs incidentally is possible, which is in line with the above-cited findings of Hare (2010) and Shin and Christianson (2012).

Magnusson and Graham (2011) compared ESL learners who received exposure-only instruction of phrasal verbs to those who received explicit instruction. Explicit instruction consisted of class discussions and visual representations of targeted phrasal verbs as well as the provision of example sentences. Fifty-five intermediate-level students completed a pretest and postest that measured both receptive and productive knowledge of 37 targeted idiomatic (i.e., opaque) phrasal verbs, which were pre-selected by four experts. The number of exposures for each phrasal verb in the learning materials was at least seven occurrences. The results show that there was a significant difference in treatment conditions. Magnusson and Graham concluded that learners learned significantly more in the explicit instruction condition, although there was some learning in the exposure-only condition. The latter result confirms again that L2 acquisition of phrasal verbs can take place incidentally.

Another study that compared the effectiveness of different instructional conditions is the study by Khatib and Ghannadi (2011). Specifically, the researchers pitted interventionist approaches to learning (both recognition and production) of phrasal verbs against noninterventionist approaches. Sixty-three EFL learners were grouped into a control group and two experimental groups. The control or *noninterventional* group read texts containing 40 targeted phrasal verbs that were selected from commercial learning materials (McCarthy & O'Dell, 2004; Watcyn-Jones, 2001), and answered post-reading comprehension questions. The experimental groups followed the same procedure, except that they were provided with marginal glosses (i.e., *input enhancement* technique), and received a brief explanation about phrasal verbs prior to reading. These groups were further divided into an *implicit interventional* group and an *explicit interventional* group. The only difference between the two being that learners in the latter group also completed post-reading production tasks. Results indicated that both experimental groups outperformed the noninterventionist (incidental) group in both recognition and production of phrasal verbs. At the same time, the interventional explicit outperformed the interventional implicit group.

The results from the studies cited above by Magnusson and Graham (2011) and Khatib and Ghannadi (2011) indicate that exposure alone is not sufficient for L2 phrasal verb learning to occur, and thus engaging learners' attention in some way is needed to improve learning outcomes (Robinson et al., 2012; Schmidt, 2001; Sharwood Smith, 1991, 1993). Therefore, similar to Khatib and Ghannadi, some researchers have looked at whether input enhancement can render phrasal verbs more salient in the hope that will be *noticed* by learners (see Sharwood Smith, 1991, 1993). Specifically, they have investigated whether textual modifications, such as **bolding**, can be effective in attracting learners' attention to targeted phrasal verbs, and hence increase the likelihood that learning will occur. As an example, Birjandi et al. (2015) carried out a study to examine the effectiveness of three types of input, including unenhanced input, typographically enhanced input and lexically elaborated input on learning English phrasal verbs by intermediate EFL learners. Thirty phrasal verbs from McCarthy and O'Dell (2004), a corpus-based reference and practice book, were used as targets. After taking a pretest assessing receptive knowledge of these target items, 35 students belonging to an intact class were exposed to instructional targets through reading during three consecutive treatment sessions (10 targets per session). In the first session, participants read two short texts without any enhancement, took the first postest, and then wrote a summary of each text. Learners were also provided with a glossary with the definitions of the targeted phrasal verbs dealt with during the reading part of the session. The second and third sessions were similar to the first session, except for the fact in the second session the target items were highlighted in bold in the texts, and in the last session the texts contained redundancies and parenthical definitions and explanations. Additionally, in the latter session, no glossary was given. Results showed that there was a facilitative effect of enhanced input when compared with unenhanced input in the learning of phrasal verbs, although the effect was not statistical significant; lexically elaborated input was the most effective of all.

Chiang (2013) also studied the effectiveness of textual input enhancement and lexical elaboration in the acquisition of phrasal verb in incidental reading settings. A large number of L2 English learners, 298 pre-intermediate junior-college students, participated in a three-month experiment. Learners were exposed to 10 target phrasal verbs via texts that contained from one to five occurrences of these verbs, which were textually enhanced through bolding. Lexical elaboration was implemented explicitly and implicitly. The explicit kind consisted of providing a synonym or an definition-type explanation after each target item, whereas the implicit kind was realized through a post-target appositive (renaming) explanation only. In addition, the researcher also looked at the the effect of exposure frequency. As Birjandi et al., Chiang also reported that both textual input enhancement and lexical elaboration were effective in improving phrasal verb acquisition, being most effective when they were combined. Interestingly, the positive effects of the instructional treatment were concerning with form, not meaning. Regarding exposure frequency, the results indicated that more exposure led to more learning, with four to five repeated encounters as a threshold for learning to occur.

Using another typical instructed SLA-rooted framework, *processing instruction* (VanPatten & Cadierno, 1993), Thibeau (1999) investigated phrasal verb syntax acquisition. More specifically, Thibeau compared traditional versus processing instruction. Processing instruction is an intervention that seeks to modify learners' basic processing strategies (VanPatten, 2012, p. 275). Seventy-one ESL intermediate learners participated in the study. Those in the traditional instruction group (controls) were given explicit explanations only, whereas those in the processing instruction group (experimentals) received a) grammatical explanations, b) information about processing strategies, and c) instruction in the form of the so-called *structured input activities*. The latter are input-based activities that aim to facilitate form-meaning connections (Wong, 2004). In Thibeau's (1999) study these activities included matching meaningful pictures in diagrams, answering yes/no and multiple-choice questions, completing sentences, and written narration. Processing instruction was found to result in significant learning gains in receptive syntactic phrasal verb knowledge as compared with traditional instruction.

Another group of research work has combined processing instruction and textual input enhancement in the teaching and learning of English phrasal verbs. Such is the case of a study conducted by Chiang (2012), which compared these two input-based instructional methods with traditional ones (i.e., list memorization and output oriented drills and exercises). Forty EFL low/pre-intermediate students took part in a three-month experiment. Twenty-five of the most frequent phrasal verbs from the corpora of the International Corpus of Learner English (ICLE) served as learning targets. Textual input enhancement was achieved via bolding. The combination of input-based instruction in the form of processing instructional and textual input enhancement proved to be significantly effective in assisting students in learning phrasal verbs.

Concordance-based or DDL instruction has also been employed for the acquisition of L2 phrasal verbs. Under this corpus-based pedagogical approach, learners are exposed to samples of words in concordance lines (i.e., sentences) derived from a corpus (see Gavioli, 2005; McEnery & Hardie, 2011; see also Boulton & Cobb, 2017, for a recent meta-analysis of DDL studies). A study examining this type of instruction was carried out by Azzaro (2012), who compared the pretest and posttest performance of two groups of learners who either worked with concordance or dictionary-based learning materials. Ten highly frequently occurring phrasal verbs originated from Gardner & Davies' (2007) analysis of the British National Corpus (BNC) were selected as instructional targets. Azzaro reported that DDL demonstrated to be effective in teaching phrasal verbs. Similarly, Sarab and Kardoust (2014) compared DDL and dictionary use in phrasal verb instruction and obtained analogous positive results. However, a recent study by Behzadian (2016) found that input enhancement was superior to concordance-based instruction in increasing phrasal verbs learning.

Other researchers have used other pedagogical approaches to enhance L2 phrasal verb teaching and learning. Nassaji and Tian (2010), for instance, explored the effects of individual and collaborative output tasks on phrasal verb learning. Two intact low-intermediate ESL classrooms completed reconstruction cloze tasks and reconstruction editing tasks in pairs and individually. The 16 phrasal verbs used as instructional targets were recommended by the teacher in charge of the two classrooms, and introduced prior to the treatment through an mini input-based lesson. Learning was identified by the success of task completion and the results of a pretest and a posttest. Results showed that all learners increased their knowledge of English phrasal verbs.

More recently, Garnier (2016) applied an intentional, word-focused approach to the acquisition of novel phrasal verbs, specifically via rote memorization, textbook exercises, and guessing from context. The results of immediate and post-delayed tests indicate improved learning outcomes in meaning-recall and meaning-recognition, with significantly higher scores due to L2 proficiency, and general vocabulary knowledge.

In summary, the studies reviewed in this section suggest that phrasal verbs can be acquired under incidental conditions. Most studies have used input-based approaches to investigate L2 phrasal verb acquisition (e.g., Chiang, 2012; Hare, 2010; Thibeau, 1999), whereas a handful have employed output-based alternatives (e.g., Nassaji & Tian, 2010). Further, explicit instructional interventions seem to yield better learning outcomes than implicit ones (see Goo et al., 2015; Norris & Ortega, 2000, for meta-analytic work on explicit and implicit L2 instruction). It can also be observed that the selection of target items in the studies was done using different criteria, including native intuition (Hare, 2010), teachers' suggestions (Nassaji & Tian, 2010), and learning materials (Khatib & Ghannadi, 2011). Apart from the obvious use of corpus-inspired methods by DDL-based studies (e.g., Azzaro, 2012; Birjandi et al., 2015), only few studies have, however, based their selection of targets on empirical corpus-based research findings (e.g., Chiang, 2012).

In light of the review above, this dissertation focuses on incidental phrasal verb acquisition through a combination of input and output-based approaches, namely reading authentic texts online accompanied by automatically generated multiple-choice activities using NLP technology. Moreover, it also addresses a possible shortcoming in the literature by using corpus linguistics methods to select learning targets. This is important because these items represent discourse produced by native speakers, and as such are realizations of *authentic language usage* (Gablasova, Brezina, & McEnery, 2017; L.-S. Huang, 2017; Sinclair, 1997; Sun & Wang, 2003). In particular, targets in the present work were selected from a news corpus automatically collected from the Internet (see Part II).

3.1.4 Corpus-based research on phrasal verbs

Corpus-based studies on phrasal verb use can be grouped into two broad branches: 1) those targeting native use, and 2) those comparing native and learner use. Each will be discussed in turn below.

3.1.4.1 Phrasal verb native use

There has been a rather small number of studies investigating contemporary English phrasal verb use by native speakers using corpus-based approaches (e.g., Biber et al., 1999; Davies, 2009; Gardner & Davies, 2007; J.-Y. Lee, 2015; D. Liu, 2011, 2012; Trebits, 2009). Biber et. al's (1999) pioneering work, for instance, found that phrasal verbs are most frequently used in fiction and spoken English, and much less so in news and academic discourse. Gardner and Davies (2007) devised a list of 100 most frequent phrasal verbs in English, based on the analysis of the BNC. This high-frequency list was then expanded to 150 phrasal verbs by D. Liu (2011), including phrasal verbs previously reported by Biber et al. Importantly, and unlike Biber et al., and Gardner and Davies, Liu's list contains frequency counts of both American and British English. As Biber et al., D. Liu found that phrasal verbs were highly frequent in fiction and conversation domains and much less frequent in newspapers and academic English. Recently, Trebits (2009) also reported high occurrence of phrasal verbs in English. Specifically, similar to Gardner and Davies, the researcher found that at least one phrasal verb construction is likely to appear every 200 words of written text.

3.1.4.2 Phrasal verb learner (under)use

Avoidance refers to L2 learners' preference for simple, easy and familiar language forms over complicated ones, with the aim of overcoming communication problems (X. Chen & Smakman, 2016; Laufer & Eliasson, 1993). On the topic of phrasal verbs, avoidance literature reveals that learners consistently avoid using these multi-word constructions; and, instead, they opt to use one-word verbs (Dagut & Laufer, 1985; El-Dakhs, 2016; Hulstijn & Marchena, 1989; Liao & Fukuya, 2004; Sjöholm, 1995).

According to Alejo-González (2010, p. 55), however, there is a difference between the phenomenon of avoidance, which presupposes that learners do *know* the L2 form, but avoid using it because they think they cannot use it accurately, and *underuse*, which can be described as the nonnatives' tendency to use, on average, fewer phrasal verbs than natives. Focusing on the latter, corpus-based studies are set out to measure learners' underuse of phrasal verbs by making the analysis of L2 learner corpora and its comparison with native corpora (e.g., Alejo-González, 2010, 2012; M. Chen, 2013a, 2013b; Deshors, 2016; Fadanelli, 2012; Gilquin, 2015; Kamarudin, 2013; Ke, 2014; Rahman & Abid, 2014; Rong, 2015; Rosca & de Altamirano, 2016; Ryoo, 2013; Siyanova & Schmitt, 2007; Sung, 2017; Sung & Kim, 2016; Uchida, 2012; Waibel, 2007; Yoshitomi, 2006). Crucially, corpus research findings are in line with those of the avoidance studies in that L2 learners produce fewer phrasal verbs in their discourse, and in that they present a distinct preference for one-word synonyms. For instance, Waibel (2007) found that Italian advanced learners of English used a much lower number of phrasal verbs in their texts than native speakers. However, Waibel also reported that L1 German students overused phrasal verbs compared to native students. Siyanova and Schmitt (2007) showed that nonnatives were less likely to use phrasal verbs than natives, and again found that learners show a stronger preference for one-word verbs. Alejo-González (2010) reported that English speakers use more than twice as many phrasal verbs with the particle *out* when compared to learners. Gilquin (2015) obtained similar results as natives' overall frequency counts of phrasal verbs significantly doubled that of students. M. Chen (2013a) observed that Chinese university students presented much lower overall frequencies of phrasal verbs than their American peers. Most recent work by Sung (2017) reported that Korean learners of English underuse phrasal verbs, and that they have a greater preference for one-word verbs over phrasal verbs.

Corpus research also indicates that learners are prone to produce more semantically transparent phrasal verbs, whereas native speakers are more likely to use idiomatic (i.e., opaque) ones (e.g., Rahman & Abid, 2014; Uchida, 2012; Yoshitomi, 2006). Another interesting finding is the fact that the frequency of phrasal verbs in learners' productions gradually increases with proficiency (Rahman & Abid, 2014; Uchida, 2012). Overall, learners use fewer phrasal verbs than native speakers.

3.1.5 L2 knowledge of phrasal verbs

Based on L2 vocabulary research, *receptive* knowledge of a phrasal verb can be understood as the learner's ability to recognize this item when encountered while reading or listening, whereas *productive* knowledge can be thought as the ability to use a phrasal verb in speaking and writing (see González Fernández & Schmitt, 2017; Nation, 2001; Read, 2000).

There is only a dearth of research that has directly measured the degree to which L2 learners know phrasal verbs receptively and productively (e.g., Cornell, 1985; Garnier & Schmitt, 2016; Hare, 2010; Rahman & Abid, 2014; Schmitt & Redwood, 2011; Shin & Christianson, 2012). For instance, Schmitt and Redwood (2011) found that learners recognized the majority of the targeted phrasal verbs (65.2%), and produced about half of them (48.2%). Garnier and Schmitt (2016) too gauged phrasal verb productive mastery, but unlike Schmitt and Redwood, who tested the most typical meaning sense, they investigated the production of several meaning senses as well. When considering the most common meaning sense, the results showed that learners used the target phrasal verbs productively at a lower rate (44.5%) than learners in Schmitt and Redwood's study. When taking into account all the meaning senses as a whole (i.e., 100), learners were able to produce around 40% of these. More generally, the findings from the above-discussed studies underscore the difficulties learners undergo in using English phrasal verbs.

3.1.6 Summary of the research review on phrasal verbs

In this first part of the chapter, the criterion assumed in this dissertation to define English phrasal verbs was explained. This criterion is based on syntactic considerations, since the detection and extraction of phrasal verbs from a news corpus was carried out automatically (i.e., computer-assisted). Further, the different difficulties learners face when learning phrasal verbs were described. These difficulties originate from semantic, syntactic and pragmatic intricacies, with the semantic dimension being perhaps the most difficult one. Moreover, the relevant literature on L2 phrasal verb acquisition, with particular emphasis on studies within the field of instructed SLA, was reviewed. This review indicates that phrasal verbs are susceptible to incidental learning, and that more explicit instructional treatments have proven to be more effective in the teaching of these multi-word expressions.

With respect to corpus-based research, the analysis of native corpora has shown that phrasal verbs are highly frequent constructions in English, which mostly occur in informal registers, and, to a lesser degree, in formal written discourse such as news texts. When comparing between native and nonnative use of phrasal verbs, corpus work has indicated that the latter typically use fewer phrasal verbs, with a marked predilection for one-word verb forms. Furthermore, the scarce literature on learners' knowledge of phrasal verbs in English has found that productive knowledge is limited, whereas receptive knowledge is broader, which is in line with research stating that recognizing language forms is easier than producing them (Daskalovska, 2011; Laufer, 2012; Melka, 1997; Schmitt, 2010; Webb, 2008). The next part describes a corpus-based study that aimed to automatically collect phrasal verb targets and experimental learning materials for the current study.

3.2 Part II: Corpus-based study

3.2.1 Methodology

3.2.1.1 Automatic construction of a news corpus

As mentioned earlier, a general corpus of news texts was automatically constructed from the Internet with the assistance of NLP software, such as the Python[7] libraries *Beautiful Soup* (https://www.crummy.com/software/BeautifulSoup/) for web scraping (i.e., automatic web data extraction), and Natural Language Tool Kit (NLTK, http://www.nltk.org/) for text preprocessing, respectively. Automatic construction refers here to the creation of a corpus by a computer program (see Hassel, 2001, 2007). To this end, a series of steps were undertaken.

Procedure. First, an online survey was set up to collect reading topics that were later used to automatically crawl (i.e., browse) the web for news texts. The survey was specific to Reuters News texts, and involved asking English L2 learners to conduct ten searches about topics they would be interested in reading in this news site. To complete the survey, participants were sent a link (http://purl.org/icall/werti-topics). The survey results yielded 139 unique reading topics, with the term "refugee crisis" with highest number of instances (4). The next step consisted of updating the existing Python code implemented in the WERTi system (Meurers et al., 2010), in order to search and retrieve news from Reuters containing the reading topics gathered from the survey. More specifically, the code automatically browsed the Reuters Rich Site Summary (RSS) feed, a web document listing the latest news from Reuters, and then automatically downloaded the texts containing the search word. Additional modifications to the code included filtering out news texts with less than 100 words as well as some preprocessing of the texts (e.g., punctuation fixes) for further parsing. The resulting nontagged corpus consisted of 11,747 texts, nearly 10,000,000 tokens[8] in total.

[7]Python is a general purpose object-oriented programming language (www.python.org)

[8]A *token* is a sequence of characters, or basic units such as words and punctuation found in a particular document (Bird, Klein, & Loper, 2009, p. 7).

3.2.1.2 The target corpus

The target corpus was a subpart of the above-mentioned corpus containing 40 written news texts from Reuters (totaling over 49,000 tokens). The average length of the texts was approximately 1,225 tokens (≈ 2 pages each). The texts covered a variety of topics including economy, technology, world news, interviews, health, entertainment, sports, and lifestyle news. The selection of the texts was based on the number of the phrasal verbs in the news text and the potential of the news item to become outdated in a short period of time. For instance, a relatively high number of instances of phrasal verbs were found in news articles reporting sports events. However, these articles were not considered because they became obsolete very quickly.

3.2.1.3 Automatic detection and extraction of phrasal verbs

As pointed out already, the syntactic dependency parser from spaCy was used to automatically detect and extract phrasal verbs from the target corpus. To reiterate, the dependency parser identifies relations among words, and of interest is the relation between the verb and the particle. In order to carry out the task at hand, a Python program was developed. This program parsed (i.e., syntactically analyzed) every text in the target corpus, sentence by sentence, as demonstrated in Figure 2.

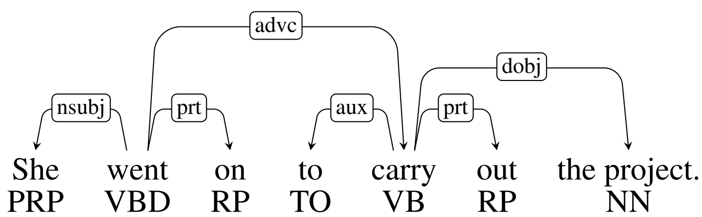


Figure 3.2. Dependency-annotated sentence containing the phrasal verbs *went on*, and *carry out*.

After the parsing was completed, the program extracted every instance of a phrasal verb it encountered (i.e., every token labeled as a particle [*prt*] and headed by a verb) and built a report of phrasal verb occurrences in the target corpus, which also included the lemmatized (i.e., no morphologically inflected) version of the detected phrasal verb. For instance, the parsed sentence in Figure 2 would produce two instances of phrasal verbs: *went on* and *carry out*. We found that the program had a precision of .95 (i.e., 95% of the spotted phrasal verbs were actually such), and a recall of .83 (i.e., 83% of actual phrasal verbs were detected). False positives were later manually unenhanced, i.e., no multiple-choice selections were automatically generated when they appeared in the target texts for the form-focused group (see Chapter 5).

3.2.2 Results and discussion

3.2.2.1 Overall frequencies of phrasal verbs in the target corpus

On the whole, 165 distinct phrasal verbs were identified in the target corpus. Likewise, there were 296 occurrences of phrasal verbs in total. When the frequency is normalized (see Evison, 2010), that is, the raw frequency of occurrences (296) is divided by the total token count of the target corpus (49,000), and then multiplied by a base number (1,000), a phrasal verb was likely to occur about six times per thousand tokens in the 40 news texts. To put it differently, on average, one phrasal verb roughly occurred every 200 tokens, which is in line with previous research that has found a similar frequency proportion (e.g., Gardner & Davies, 2007; Trebits, 2009).

3.2.2.2 Target phrasal verbs

Of the overall number of phrasal verbs in the target corpus (165), a subset of 30 (18.2%) was chosen to serve as instructional targets. The selection was based on frequency of occurrence, with each phrasal verb appearing at least twice in the 40-text corpus. Table 3.1 present a list of these phrasal verb constructions. Concerning general frequency, this table also shows that 21 (70%) of the learning targets are included in Liu's (2011) list of 150 most frequent phrasal verbs in American and British English (i.e., at least 10 tokens per million words; rank in parenthesis), which is the most recent corpus-informed analysis of frequency of English phrasal verbs. This list has also been used as reference in recent L2 literature on phrasal verbs (e.g., Garnier & Schmitt, 2015, 2016). In other words, most instructional targets in this dissertation are highly occurring items in English.

Table 3.1

Frequency of target phrasal verbs in the subcorpus

Rank	PV	Raw frequency	% of total PVs
1	carry out (36)	9	5.45
2	go on (1)	8	4.85
3	take on (15)	8	4.85
4	set up (11)	7	4.24
5	take over (37)	7	4.24
6	end up (18)	6	3.64
7	open up	6	3.64
8	give up (16)	5	3.03
9	grow up (10)	5	3.03
10	make up (17)	5	3.03
11	set out (64)	5	3.03
12	blow up (99)	4	2.42
13	clean up (65)	4	2.42
14	come up (4)	4	2.42
15	pay off (78)	4	2.42
16	sign up	4	2.42
17	buy up	3	1.82
18	figure out (21)	3	1.82
19	lay off	3	1.82
20	lay out (73)	3	1.82
21	point out (9)	3	1.82
22	rule out (123)	3	1.82
23	scale up	3	1.82
24	shore up	3	1.82
25	shut down (66)	3	1.82
26	water down	3	1.82
27	build up (84)	2	1.21
28	gobble up	2	1.21
29	ramp up	2	1.21
30	turn out (12)	2	1.21

Note. PV = phrasal verb. Numbers in parentheses indicate the frequency ranking order in the list of most frequent phrasal verbs in English (D. Liu, 2011).

Table 3.2 displays the semantic analysis of the target phrasal verbs. As noted previously, phrasal verbs are highly polysemous (Gardner & Davies, 2007; D. Liu, 2011). Therefore, it is important to determine the meaning senses of target phrasal verbs in order to assess them in both the pretest and the posttest. On the basis of the PHaVE List, which contains the most frequent meaning senses of the most frequent phrasal verbs in English (Garnier & Schmitt, 2015), and other resources such as various online dictionaries, a nonautomatic semantic analysis of the target phrasal verbs was performed. Overall, there were 37 meaning senses, with most of the target phrasal verbs (83.3%) displaying one meaning sense. In the case of those phrasal verbs with more than one meaning sense, only the most frequent one was assessed. Given that learners tend to use Latinate or one-word synonyms instead of phrasal verbs (Liao & Fukuya, 2004; Sung, 2017), as earlier stated, one-word verbs were used as near synonyms in the tests (see Chapter 5).

Relatedly, regarding semantic opacity, the meaning of a phrasal verb can be literal, and figurative and/or idiomatic (Garnier & Schmitt, 2016; Ke, 2016). For Ke (2016), phrasal verbs with a literal meaning sense are those in which the verb tends to maintain its original meaning and the particle either indicates direction (adverbial function), or contributes with little or no meaning (e.g., *go out*). In phrasal verbs with figurative meaning senses, on the other hand, the combination verb plus particle has a metaphorical (i.e., symbolic) meaning, and/or the meaning cannot be deduced by the sum of its parts (e.g., *turn up* [arrive]). Distinguishing between literal and figurative phrasal verb meaning senses is not a trouble-free task, since one construction can sometimes belong to either semantic category (Rodríguez-Puente, 2012). Consequently, in this dissertation, these labels were assigned by consulting with native speakers of English, which resulted in most phrasal verbs being labeled as *figurative* (63.3%) (see table 3.2).

Using NLP to both collect and analyze the target texts automatically allowed the researcher to address a potential methodological weakness in the studies on phrasal verb acquisition, namely using a corpus-based criteria to select learning targets and materials. As noted earlier, this selection is seldom done on the basis of an actual corpus analysis, such as the one carried out here. However, one of the challenges of using authentic texts (i.e., texts intended for native speakers, not learners) for experimental L2 research is that the target form may not be frequent enough in these materials. In the case of phrasal verbs, as reviewed above, they are to be fairly frequent in English. Nonetheless, it was found that, as reported in the corpus literature (e.g., Biber et al., 1999; D. Liu, 2011), they were less frequent in news texts. Consequently, it was difficult to build an online news corpus in which distinct phrasal verbs appeared more than nine times at most across the 40 news texts (see Table 3.1).

Table 3.2

Semantic analysis of target phrasal verbs

PV	Meaning	Semantic category
blow up	explode	L
build up	increase	F
buy up	acquire	L
carry out	complete	F
clean up	sanitize	F
come up	produce	F
end up	finish	F
figure out	understand	L
give up	abandon	L
go on	proceed	L
gobble up	consume	F
grow up	mature	L
lay off	disemploy	F
lay out	present	F
make up	constitute	F
open up	expand	F
pay off	liquidate	L
point out	indicate	F
ramp up	intensify	F
rule out	exclude	F
scale up	maximize	L
set out	present	F
set up	establish	L
shore up	support	F
shut down	close	F
sign up	register	L
take on	accept	F
take over	assume	L
turn out	prove	F
water down	attenuate	F

Note. L = Literal. F = Figurative

3.2.3 Summary of the corpus-based study

The second part of the chapter reported the results of a corpus-based study that allowed the selection of learning materials and phrasal verb targets. The corpus collection and syntactic analysis was done automatically using NLP. The analysis yielded a target corpus of 40 news texts containing 165 unique phrasal verbs, 30 of which were chosen to be the learning targets. Each targeted phrasal verbs appeared at least twice throughout the entire 40-text corpus. A manual semantic analysis identified 30 one-word phrasal verb meanings to be assessed, which were predominantly figurative in nature.

3.3 Summary

This chapter concerned the selection of linguistic target and comprised two parts. In the first part, it presented a review of literature on the acquisition of English phrasal verbs. In the second part, it described a corpus-based study that included both the automatic collection and analysis of a news corpus from the Internet. In the next chapter, a study comparing performance between lab-based and web-based versions of working and declarative memory tests is reported. The online versions of these cognitive were used in the primary experiment of the dissertation (see Chapter 5).

Chapter 4

Comparing lab- and web-based performance in individual difference measures

This chapter reports the results of a study that sought to validate web-based versions of cognitive tests to be employed in the main study of the present dissertation. More specifically, it examined whether lab-based and web-based versions of working and declarative memory tests yielded similar performance scores, i.e, whether the two versions were equivalent. The chapter first briefly explains why individual differences, particularly working and declarative memory learning abilities, were investigated here. It also justifies the selection of tests to measure these constructs. This is followed by the description of the study comparing performance in the lab-based and web-based versions of the tests, detailing methodology, results, and discussion of the findings.

4.1 Individual differences

Individual differences are a central factor in learning and instruction (Jonassen & Grabowski, 1993). In L2 learning, learners' cognitive abilities have been shown to have great explanatory power in accounting for differences in learning outcomes (Dörnyei, 2003, 2005). Among these, working memory and declarative memory are deemed to be very important sources of learners' variability (Grey et al., 2015; Hamrick, 2015; Lado, 2017; Li, 2015, 2017a, 2017b; Morgan-Short et al., 2014; Tagarelli et al., 2015, 2016).

Measuring cognitive abilities is usually performed in a lab-based setting where data collection can be an expensive and laborious process. This may affect the study's sample size and statistical power, and thus limiting the generalizability of findings (Krantz & Reips, 2017; Plonsky, 2017). Web-based testing could contribute to reducing these concerns as, for example, a larger number of participants can be tested fast, and administration is relatively easier and more cost-effective than research conducted in the lab (Gelman, 2018; Klaus & Schriefers, 2016; Reips & Birnbaum, 2011; Shrout & Rodgers, 2018; Stewart, Chandler, & Paolacci, 2017). Moreover, web-based experimenting has been found to be a reliable and effective research tool (Hicks et al., 2016; Jylkkä et al., 2017; Klaus & Schriefers, 2016; Linnman, Carlbring, Åhman, Andersson, & Andersson, 2006; Reips, Buchanan, Krantz, & McGraw, 2015; Semmelmann & Weigelt, 2017).

Researchers comparing web-based and lab-based experiments on psychological constructs appear to agree that the two modes yield comparable results (Krantz, Ballard, & Scher, 1997). However, most web-based research has been typically done through surveys (Wolfe, 2017), including those examining individual differences (e.g., Horzum, İsmail Önder, & Şenol Beşoluk, 2014; Zettler, 2011). Relatively fewer attempts have made to compare performance produced by lab-based and web-based versions of cognitive tests (Hicks et al., 2016; Jylkkä et al., 2017; Klaus & Schriefers, 2016; Linnman et al., 2006; Semmelmann & Weigelt, 2017). One of these attempts, and of particular relevance here, was carried out by Hicks et al. (2016), who correlated scores obtained in both lab-based and web-based versions of different working memory tests. In a series of experiments, Hicks et al. sought to replicate the association between working memory and fluid intelligence in a web-based testing environment. After the first two experiments, the researchers determined that participants had cheated in the verbal working memory tests (i.e., operation span, reading span, and running span) by writing down the letter memoranda. As a result, Hicks et al. designed two tasks in which the material to be remembered was more abstract in nature (i.e., pictures of objects and Klingon symbols), thereby increasing the validity of the web-based tests. As a whole, the researchers found evidence that performance in the two settings was comparable or equivalent in all measures, i.e., there was a consistent pattern of correlations between working memory and fluid intelligence across both test modes. An adaptation of the Klingon Span Task was used in the current study (see Methods).

Equivalence between two measurements can be of two types (American Psychological Association, 1986; Gwaltney, Shields, & Shiffman, 2008; Schulenberg & Yutrzenka, 1999). The first type concerns whether two measurements yield the same numbers. The second type of equivalence is related to whether differences found in one measurement are also systematically found in the other. This means that, although the two measurements estimate two different numbers, these numbers have a systematic and very clear relationship to each other.

This chapter presents the results of a study that investigated the differential performance generated by two versions of tests measuring working memory and declarative memory capacities in lab-based and web-based settings, with the aim to determine whether the two versions are equivalent with respect to the second type of measurement equivalence, i.e., whether the scores in the two versions correlated. Establishing measurement equivalence between these two administration modes is crucial for several reasons. First, it is necessary to show that the results of web-based studies are comparable to those of previous research, which have predominantly obtained from data gathered in lab-based settings. Second, it is imperative to ensure that cognitive constructs are measured in the same way in both test modes. Finally, it is important to ascertain whether lab-based and web-based measures are equivalent because, if they are, web-based testing could be a feasible alternative to address some of the current methodological issues found in research conducted in lab-based settings, such as restricted population sampling, underpowered studies, and small sample sizes, among others (S. K. T. Bailey, Neigel, Dhanani, & Sims, 2017; Capielo, Mann, Nevels, & Delgado-Romero, 2014; Gelman, 2018; Gwaltney et al., 2008; Leeson, 2006; McDonald, 2002; Schulenberg & Yutrzenka, 1999; Shrout & Rodgers, 2018; Stewart et al., 2017; Vandenberg & Lance, 2000).

4.1.1 Working memory

Working memory refers to the capacity to simultaneously process and retain information while carrying out complex cognitive tasks such as language learning, comprehension and production (Cowan, 2016; Peng et al., 2017). Following Baddeley and colleagues (e.g., Baddeley, 2000; Baddeley & Hitch, 1974), working memory is a multicomponent system that consists of storage subsystems that are responsible for holding visualspatial and auditory information, an episodic buffer that acts as a link between the storage subsystems and long-term memory, and a central executive that functions as an attentional control system.

In L2 learning, working memory is believed to assist learners in jointly processing form, meaning and use of language forms at the same time (Doughty, 2001; J. N. Williams, 2012). Its potential role to account for individual differences in L2 outcomes has been indicated by meta-analytic work (Grundy & Timmer, 2017; Linck et al., 2014), and such a role has been examined in different areas, including speaking (e.g., O'Brien et al., 2007; Weissheimer & Borges Mota, 2009), vocabulary (e.g., Hong & MacWhinney, 2011; Hummel, 2009; Kempe et al., 2009; Martin & N. C. Ellis, 2012; Yang et al., 2017) and reading (e.g., Alptekin et al., 2014; Medina et al., 2017; Sagarra, 2017). Most research, however, has been done on the area of syntax (e.g., Indrarathne & Kormos, 2018; Tagarelli et al., 2015, 2016).

Working memory has been measured by simple and complex span tasks. Simple span tasks (e.g., digit span and letter span) involve recalling short lists of items, and they seek to gauge the storage aspect of working memory (H. Bailey et al., 2011). Complex span tasks, such as the operation span task (OSpan; Turner & Engle, 1989), on the other hand, entail remembering stimuli while performing a secondary task, and are thought to tax both processing (attention) and storage (memory) components of working memory (Baddeley, 2000; H. Bailey et al., 2011).

4.1.2 Declarative memory

Declarative memory is understood as the capacity to consciously recall and use information (Reber et al., 1996). It is posited to depend on the declarative memory system, one of the long-term systems in the brain responsible for the processing, storage, and retrieval of information about facts (semantic knowledge) and events (episodic knowledge; Eichenbaum, 2004; Squire, 2004). Learning occurring under this system is to be quick, intentional, attention-driven, and predominantly explicit (M. M. Chun, 2000; Knowlton et al., 2017; Ullman & Lovelett, 2018).

Moreover, learning through the declarative memory system is suggested to result after a single exposure of a stimulus, although it is strengthened via increased exposures (Lum et al., 2012; Ullman & Lovelett, 2018).

Declarative memory is assumed to play a role, for example, in first and second language acquisition (Morgan-Short et al., 2014). In the latter, it is thought to underlie the learning, storage and processing of L2 vocabulary and grammar (Paradis, 2004, 2009; Ullman, 2015, 2016), at least in the earliest phases of acquisition (Hamrick et al., 2018; Morgan-Short et al., 2014). The predictive ability of declarative memory to explain variation in L2 attainment has been attested in several studies (e.g., Carpenter, 2008; Faretta-Stutenberg & Morgan-Short, 2018; Hamrick, 2015; Morgan-Short et al., 2014).

Declarative memory capacity can be gauged via recall and recognition tasks (Morgan-Short et al., 2014), both verbal, such as the paired associates subtest of the Modern Language Aptitude Test (MLAT5; J. B. Carroll & Sapon, 1959), and nonverbal, such as the Continuous Visual Memory Task (CVMT; Trahan & Larrabee, 1988). In this dissertation, the MLAT5 and the CVMT were used to index verbal and nonverbal declarative memory, respectively. Even though these two assess different types of learning (verbal vs. visual), they both entail declarative memory processes, i.e., memory of facts and/or events (Buffington & Morgan-Short, 2018; Eichenbaum, 2011). Moreover, both the MLAT5 and the CVMT have been employed in SLA studies before (e.g., Carpenter, 2008; Faretta-Stutenberg & Morgan-Short, 2018; Morgan-Short et al., 2014).

4.2 Comparison study

As noted earlier, the comparison study aimed to validate web-based versions of working memory and declarative memory tests that were to be used the main study. Specifically, it sought to establish whether lab-based and web-based versions produced similar performance scores, i.e, whether the two versions were equivalent. To that end, it assessed whether the values of one type of mode of administration corresponded to the values in the other mode (i.e., second type of equivalence). Put simply, are the differences in scores constant, or parallel in the two ways of measuring?

The implications of this study for future research using these tests as variables are as follows. If lab-based and web-based versions are equivalent in the second sense of equivalence, the lab-based and web-based measurements cannot be mixed, but if researchers use one version exclusively throughout a study, they will obtain the same results as if they exclusively use the other version. If the two versions are not equivalent, then they might produce different results that are dependent on the test mode where they are implemented.

4.3 Methods

4.3.1 Participants

Fifty participants (37 women and 13 men), with a mean age of 26.4 years (*SD* = 4.2), took part in the study. Most participants were native speakers of German (72%), followed by Russian (8%), Spanish (6%), Chinese (4%), English, Hungarian, Persian, Serbian and Vietnamese (2% each). Seven (14%) participants did not complete the second half of the study (i.e., web-based testing). All participants gave informed consent and received €20 for participating.

4.3.2 Materials

Three cognitive tests were administered, one assessing working memory capacity, and two indexing verbal and nonverbal declarative memory capacity, respectively. In the lab-based context, working memory and nonverbal declarative memory tests were programmed and delivered via E-Prime v2.0 (Schneider, Eschman, & Zuccolotto, 2002); the verbal declarative memory test was applied in paper-pencil form. For the web-based mode, versions of the three cognitive tests were developed for this study using Java with the Google Web Toolkit (`http://www.gwtproject.org`), and were accessible from all browsers. The tests are described below.

4.3.2.1 Working memory

To assess participants' working memory capacity, an adapted version of the Automated Operation Span Task (OSpan; Unsworth et al., 2005), a computerized form of the complex span task created by Turner and Engle (1989), was used (Baranski & Was, 2018; Faretta-Stutenberg & Morgan-Short, 2018; Hicks et al., 2016; Jordano & Touron, 2017). This adaptation was based on the Klingon Span Task developed by Hicks et al. (2016), and consisted of replacing letters, the original stimuli to be remembered in the OSpan task, with Klingon symbols. As indicated earlier, Hicks et al. implemented this change because their research showed that participants were cheating by writing down the letter memoranda in the web-based version of the classic OSpan.

The task took approximately 25 minutes to complete, and was divided into a practice phase and a testing phase. In the practice phase, participants were first presented with a series of Klingon symbols on the screen, and were asked to remember them in the order they had appeared at the end of each trial (i.e., symbol recall). Next, participants were asked to solve a series of simple math operations (e.g., $5 * 2 + 1 =?$). Finally, participants performed the symbol recall while also solving the math problems, as they would do later in the actual testing phase. After the practice phase, participants were presented with the real trials, which consisted of a list of 15 sets of 3–7 randomized symbols that appeared intermixed with the equations, totaling 75 symbols and 75 math problems. At the end of each set, participants were asked to recall the symbols in the sequence they had been shown. An individual time limit to answer the math problems in the real trials was derived from the average response time plus 2.5 standard deviations taken during the math practice section. Following Unsworth et al. (2005), a partial score (i.e., total number of correct symbols recalled in the correct order) was taken as the OSpan score (cf. Conway et al., 2005, for a description of scoring procedures). The highest possible score was 75.

4.3.2.2 Verbal declarative memory

The Modern Language Aptitude Test, Part 5, Paired Associates, (MLAT5; J. B. Carroll & Sapon, 1959) was used as a verbal measure of declarative memory (Carpenter, 2008; Faretta-

Stutenberg & Morgan-Short, 2018; Morgan-Short et al., 2014). The MLAT5 required participants to memorize artificial, pseudo-Kurdish words and their meanings in English. Participants first studied 24-word association pairs for two minutes, and then completed a two-minute practice section. During the practice section, the list of foreign words and their English equivalents were made available for participants to refer back if they needed to. Finally, participants completed a timed multiple-choice test (four minutes), in which they were asked to select the English meaning of each of the 24 pseudo-Kurdish words from five options previously seen at the memorization stage. For each correct response, one point was awarded, yielding a total score of 24 points. The test duration was 8 minutes.

4.3.2.3 Nonverbal declarative memory

The Continuous Visual Memory Task (CVMT; Trahan & Larrabee, 1988) was included as an assessment of nonverbal declarative memory (Carpenter, 2008; Faretta-Stutenberg & Morgan-Short, 2018; Morgan-Short et al., 2014). The CVMT is a visual recognition test that involves asking participants to first view a collection of complex abstract designs on the screen and then to indicate whether the image they just saw was novel ("new") in the collection, or they had seen the image before ("old"). Seven of the designs were "old" (target items), and 63 were "new" (distractors). Throughout the task, the target items appeared seven times (49 trials), and the distractors only once (63 trials). All items were presented in a random but fixed order, each one appearing for two seconds. After the two seconds, participants were instructed to respond to the "OLD or NEW?" prompt on the screen. In the lab-based setting, participants indicated their choice by mouse clicking either left for "NEW", or right for "OLD". In the web-based setting, they responded by pressing either the "N" key for "NEW", or the "O" key for "OLD" on the keyboard. Overall, the CVMT required 10 minutes to be completed. For each participant, a *d'*(d-prime) score (Wickens, 2002) for CVMT was computed. The *d'* score was used to account for the possible participants' response bias toward choosing "OLD" or "NEW".

4.3.3 Procedure

As previously noted, participants completed two cognitive testing sessions, one in the lab and one on the web. In the lab-based session, in the presence of a proctor, each participant was tested individually. After providing informed consent, participants took the three cognitive tests under investigation in fixed order: OSpan, CVMT, and MLAT5. They were then asked to fill in a background questionnaire. The whole lab-based session took about 40 minutes.

For the web-based session, each participant was sent an email containing a unique web link with a personalized code, that when clicked, took them to an interface housing the web-based versions of the cognitive tests. To prevent participants from taking the tests multiple times, the link became nonfunctional once they had submitted their responses in the last test (i.e., MLAT5). In the email, participants were also informed that the web-based session lasted about 40 minutes, and had to be completed within a week. On the interface, following informed consent, participants were given general instructions in accordance with the web-based nature of the experiment. These instructions included completing the experiment in a quiet place without interruption, and from start to finish in one sitting. Participants were also instructed not to use the browser's back button, or refresh the browser page, or close the browser window. Importantly, they were told not to take any notes during the entire experiment. The detailed instructions are reproduced in the appendix. The tests were taken in the same fixed order as in the lab-based session. The mean period between the first and second testing was 45.7 days (*SD* = 4.1).

4.3.4 Data analysis

All data were analyzed using the statistical software package R version 3.3.2 (R Development Core Team, 2016). To develop linear regression models, the lm function in the lme4 package (D. Bates, Mächler, Bolker, & Walker, 2015) was employed.

4.4 Results

Descriptive statistics are presented first, followed by the results of regression analyses. From a temporal point of view, lab scores were used to predict web scores in the linear regression models. To verify normality, model residuals were visually inspected.

4.4.1 Descriptive statistics

Table 4.1 presents the descriptive statistics summarizing participants' performance on the three cognitive tests under investigation in both testing modes.

Table 4.1

Descriptive statistics for comparison of lab-based and web-based testing.

Test	Min	Max	*M*	*SD*
OSpan Lab-based	4	59	25.78	13.34
OSpan Web-based	7	75	29.79	15.42
MLAT5 Lab-based	4	24	17.92	5.50
MLAT5 Web-based	3	24	19.10	5.81
CVMT Lab-based	0.97	3.09	1.99	0.46
CVMT Web-based	1.10	3.93	2.30	0.63

Note. OSpan = Automated Operation Span Task; Verbal declarative memory test: MLAT5 = Modern Language Aptitude Test, Part 5; Nonverbal declarative memory test: CVMT = Continuous Visual Memory Task.

4.4.2 Regression analysis

The results of the regression analyses are displayed in Table 4.2. For the working memory test (OSpan), the unstandardized coefficient was .89 (β = .77, *SE* = 0.10, $p < .001$). For the verbal declarative memory test (MLAT5), the unstandardized coefficient was .83 (β = .78, *SE* = 0.09, $p < .001$). And for the nonverbal declarative memory test (CVMT), the unstandardized coefficient was .74 (β = .54, *SE* = 0.19, $p < .001$). Overall, the results showed that the lab-based and web-based scores are significantly correlated.

Table 4.2

Regression results for comparison of lab-based and web-based scores.

Test	Unstandardized coefficient[a]	*SE*	*p*
OSpan	0.89 (.77)	0.10	< .001
MLAT5	0.83 (.78)	0.09	< .001
CVMT	0.74 (.54)	0.19	< .001

Note. OSpan = Automated Operation Span Task; MLAT5 = Modern Language Aptitude Test, Part 5; CVMT = Continuous Visual Memory Task.
[a]The standardized coefficient (β) in parentheses.

4.5 Discussion

Estimating an individual's working memory and declarative memory capacities has been predominantly accomplished in the lab-based setting (e.g., Allen, Hitch, & Baddeley, 2018; Baranski & Was, 2018; Conti-Ramsden, Ullman, & Lum, 2015; Gonthier, Aubry, & Bourdin, 2017; Ingvalson, Nowicki, Zong, & Wong, 2017; Shipstead, Harrison, & Engle, 2016). This study was an attempt to do this on the web, and compare performance in the two settings. This is important because if web-based testing proves to be comparable to lab-based testing, researchers will be able to reach more participants for their studies, which, in turn, can help alleviate some of the current concerns in L2 research (e.g., low statistical power, nonrepresentative population samples, and small sample sizes). On the other hand, ensuring that lab-based and web-based versions are equivalent is essential for the comparability of results across studies. Crucially, establishing measurement equivalence between lab-based and web-based versions will also provide assurance that the tests are measuring cognitive constructs the same way regardless of administration mode (S. K. T. Bailey et al., 2017; Capielo et al., 2014; Gelman, 2018; Gwaltney et al., 2008; Hicks et al., 2016; Krantz & Reips, 2017; Schulenberg & Yutrzenka, 1999; Shrout & Rodgers, 2018; Stewart et al., 2017). With this in mind, this study set out to examine whether the lab-based and web-based versions of three cognitive tests were equivalent with regard to whether differences in performance were constant in the two versions.

The results indicated that the scores in the lab-based and web-based versions of three cognitive tests (MLAT5, CVMT, OSpan) were equivalent in the sense that differences in performance were constant in the two versions. This suggests that participants who had relatively high values in one task also had relatively high values in the second, or the other way around. However, the strength of the association depended on the test. In both the working memory test (OSpan) and the verbal declarative memory test (MLAT5) the scores were more strongly correlated (β = .77 and β = .78, respectively); for the nonverbal declarative test (CVMT), equivalence appears to be weaker (β = .54). On the whole, the correlations reported here between lab-based and web-based scores are consistent with the assumption that both versions seem to likely measure the same cognitive construct, at least for the working memory test (OSpan) and the verbal declarative memory test (MLAT5), and, to a lesser extent, for the nonverbal declarative test (CVMT).

A possible explanation for the weaker equivalence found for the versions of the nonverbal declarative test (CVMT) is perhaps the difference in the way responses to the visual stimuli were input in the two versions. Recall that in the lab-based version, participants used left ("NEW") or right ("OLD") mouse clicking to enter their response, whereas in the web-based version, they used the keyboard ("N" and "O" keys). This modification was made to the web-based version because of technical reasons, i.e., the browser window may not register the participants' response if the cursor is not over a certain area on the page, which in itself may cause problems of missing data. It has been previously reported that participants in web-based research are prone to make errors when using the keyboard to enter their responses (Leidheiser, Branyon, Baldwin, Pak, & McLaughlin, 2015), which in this case might have affected the results of the comparison between lab-based and web-based versions of CVMT. Further studies comparing performance between the two versions may benefit from gathering data via touch input instead, which might overcome the technical difficulty of employing mouse clicking for web-based data collection reported here (see Leidheiser et al., 2015).

The study provided evidence that it is possible to measure individual differences in cognitive abilities on the web and obtain similar performance as in the lab. The lab-based and web-based versions of the three cognitive tests are equivalent. However, given that they do not perfectly correlate, the recommendation is to employ one of the two modes within one study and not to compare individual scores from one mode with scores from the other. Moreover, the extent to which the measures are equivalent varies according to the test. In this sense, it is likely that the two versions for the working memory test (OSpan) and the verbal declarative memory (MLAT5) are fairly possibly measuring the same construct, but this might not be the case for the nonverbal declarative test (CVMT), where the two modes might still plausibly measure strongly different aspects as well.

4.6 Summary

This chapter described the results of a study that compared performance scores obtained in lab-based and web-based versions of the three cognitive tests used in the main study. This research indicates that collecting experimentally controlled data on cognitive individual differences in the Internet is feasible and comparable to lab-based collection. In the next chapter, the methodology of the main study is outlined.

Part II

Main study

Chapter 5

Methods

This chapter describes the methodology used in the present dissertation. First, the participants are described. Next, the target linguistic form is explained. Then, a description of the instruments is provided. This is followed by an account of the data collection procedures. Finally, details of how the data were analyzed are discussed.

5.1 Participants

Before data collection, a power analysis based on a preliminary study with the same experimental setup (Ziegler et al., 2017) was performed. The Cohen's *F* statistic of the interaction effect between time of measurement and the condition was .36. If the effect size was indeed of that magnitude, and in order to preserve alpha .5 and power of .80, a sample size of 66 participants would then be required, 33 in each condition. If the effect size was slightly smaller (.30), and the preliminary study produced an overestimate, 92 participants would be needed in order to still maintain the same alpha error and the power. And if it was an underestimate (.40), a sample size of 54 participants would be warranted. Given this scenario and in order to be cautious, a smaller effect size (.32) was projected. As a consequence, data from 80 participants were intended to be collected, 40 in each condition.

In the actual experiment, 127 participants (91 women and 36 men), with a mean age of 24.33 years (*SD* = 3.41), took part in the study. They were randomly assigned to two experimental conditions: i) *meaning-focused* (i.e., reading; n = 63), and ii) *form-focused* (i.e., reading in addition to interacting with the text via selecting multiple-choice options; n = 64). Three additional participants were excluded for failing to follow task instructions, i.e., they finished the experiment within less than a week of the two weeks required (mean days = 2.67, SD = 0.47). Furthermore, due to technical difficulties, the number of participants completing the three cognitive tasks under investigation (see section 1.3.2) was as follows: working memory (OSpan) = 102; verbal declarative memory (MLAT5) = 92; and nonverbal declarative memory (CVMT) = 89. As such, the missing data required no special treatment, as mixed-effects regression modeling, the main statistical procedure used in this dissertation (see section 5.5), is robust against unbalanced data (Baayen, 2008; Jaeger, 2008).

The majority of the participants (99) were native speakers of German, followed by Chinese (4), Italian (4), Russian (4), Spanish (3), Lithuanian (2), Nepali (2), Polish (2), Bulgarian (1), Indonesian (1), Macedonian (1), Portuguese (1), Serbian (1), Swedish (1), and Vietnamese (1). Nine participants had a second mother tongue as well, and one had three.

English language proficiency was assessed using C-test scores obtained at the beginning of the experiment, with most learners at advanced level (n = 78, M = 98.70, SD = 8.16), followed by intermediate (n = 37, M = 76.95, SD = 5.09), and basic (n = 12, M = 55.07 , SD = 11.68; maximum score possible: 124 points). English reading ability was measured via the DIALANG test of reading proficiency, which was also administered at the start of the study, with learners falling between advanced (81) and intermediate (45) (see below for scoring details). Only one participant was at basic level, and therefore was excluded from any statistical analysis involving reading proficiency as a variable.

Meaning-focused and form-focused groups did not differ significantly across age, $t(124.46) = 0.06$, $p = .952$, and gender, $\chi^2 = (1, N = 127) = 0.86$, $p = .353$. All participants gave informed consent and received €60 Amazon e-vouchers for participation.

5.2 Linguistic target

As noted in Chapter 3, English phrasal verbs were selected as the linguistic target feature for the current dissertation given their notorious difficulty in acquisition by second language learners, even for those with an advanced level of proficiency in English (Cornell, 1985; Garnier & Schmitt, 2015, 2016; Larsen-Freeman & Celce-Murcia, 2015). For our purposes, a phrasal verb consists of a verb followed by a particle that behaves as a syntactic and lexical unit (e.g., Skype was *snapped up* [bought] by eBay Inc.). Using frequency of occurrence as a criterion, 30 out of 165 phrasal verbs derived from an automatic collected corpus of news texts from Reuters were chosen as instructional targets. The number of occurrences of the targeted phrasal verbs ranged from 2 to 9 across the 40 experimental texts (see Chapter 3, for further details).

5.3 Instruments

5.3.1 WERTi interface

The entire experiment, from the pretest stage to the posttest stage, was housed online in a customized version of the WERTi (Working with English Real Texts interactively) system, a web-based tool that aims to support language learning by providing automatic input enhancement of authentic texts (Meurers et al., 2010). Specifically, the system uses Natural Language Processing to identify targeted language items in webpages. The WERTi interface, which focuses on English, is part of the general Visual Input Enhancement of the Web platform (VIEW; `http://purl.org/icall/view`) that covers different languages and target patterns (Reynolds, Schaf, & Meurers, 2014; Ziegler et al., 2017).

The tailored adaptation of the WERTi interface employed in this dissertation supported testing procedures and instructional treatments. The interface was developed using Django, an open source web application framework written in Python (`https://www.djangoproject.com/`), which enabled the functionality to integrate, among other things, the different language and cognitive tests, participants' questionnaires as well as the treatments in one web location. The interface also supported explicit logging of all learner-system interactions during treatment based on a login that automatically assigned every participant to either the meaning-focused group or the form-focused group.

The WERTi system maintained a learner log file for each participant, recording information about the news article; the learner's interaction with the targeted forms in the form-focused condition; 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 below), the system additionally registered the total number of correct and incorrect answers in the multiple-choice gaps, 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).

The new interface also had an administrative page (i.e., "WERTi Admin") where the researcher could monitor all aspects of the experiment, from the registration phase to the culmination of the experiment. For example, Figure 5.1 shows a portion of the "Admin" page that contained information about the profile of the participants (e.g., which experimental group they had been assigned to, the number of tasks they had performed, and whether they had completed the background questionnaire).

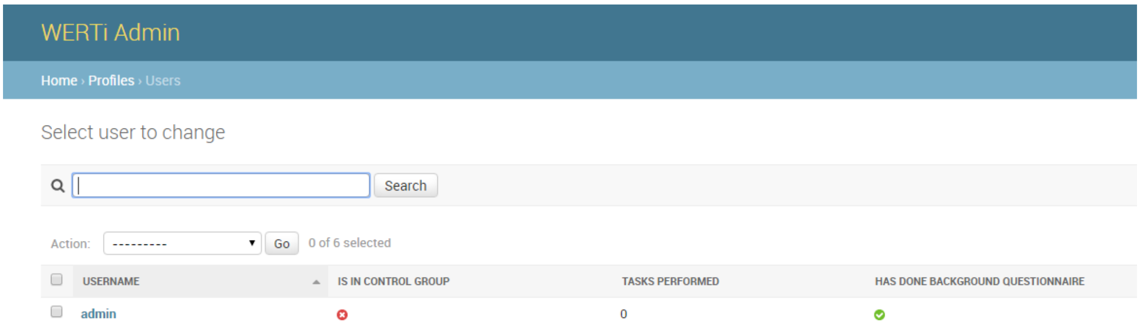


Figure 5.1. Sample of the "Admin" page.

5.3.2 Individual difference measures

As indicated in Chapter 4, in the current investigation two individual cognitive differences were examined, namely working memory and declarative memory. These cognitive constructs have been shown to influence second language learning outcomes (Grundy & Timmer, 2017; Hwu et al., 2014; Lado, 2017; Li, 2015, 2017a; Morgan-Short et al., 2014; Tagarelli et al., 2015; J. N. Williams, 2012). In the case of working memory, it is believed to have a role in assisting learners to jointly process form, meaning and use of language forms simultaneously (Doughty, 2001; J. N. Williams, 2012). Declarative memory, on the other hand, is thought to underlie the learning, storage and processing of L2 vocabulary and grammar (Ullman, 2015, 2016), at least in the first stages of acquisition (Hamrick et al., 2018; Morgan-Short et al., 2014). Particularly relevant to the present study is that both cognitive differences are seen as individual learner variables that interact with instructional conditions (Faretta-Stutenberg & Morgan-Short, 2018; Robinson, 2001a), that is, their effects may be dependent on the context where instruction occurs. In this sense, aptitude-treatment-interactions (Cronbach & Snow, 1977; Snow, 1991) have been reported for both working memory (e.g., Erlam, 2005; Faretta-Stutenberg & Morgan-Short, 2018; Goo, 2012; Indrarathne & Kormos, 2018), and declarative memory (e.g., Faretta-Stutenberg & Morgan-Short, 2018; Hwu et al., 2014; Hwu & Sun, 2012; Robinson, 1997).

In order to gauge these cognitive abilities in participants, three tests were used in the current investigation, one assessing complex working memory capacity, and two measuring verbal and nonverbal declarative capacity, respectively. Web-based versions of these tests were developed and compared with lab-based versions to determine their equivalence (see Chapter 4 for details). Moreover, to measure proficiency in English, the DIALANG reading test and the C-test were given. Descriptions of these tests are given below. In what follows, a description of each measure is presented.

5.3.2.1 Working memory

To assess participants' working capacity, a web-based adaptation of the Automated Operation Span Task (OSpan; Unsworth et al., 2005), a computerized form of the complex span task created by Turner and Engle (1989), was used. This adaptation was based on the Klingon Span task developed by Hicks et al. (2016), and consisted of replacing letters, the original stimuli to be remembered in the OSpan task, with Klingon symbols (see Chapter 4, for further details). As a dual-task measure, the OSpan taxes both processing (attention) and storage (memory) components of working memory (Baddeley, 2000).

The OSpan task took approximately 25 minutes to complete. This task was divided into a practice phase and a testing phase. The practice phase was further divided into three sections. In the first practice section, the symbol recall, participants were presented with a series of symbols on the screen, and were asked to remember the symbols in the order they had appeared at the end of each trial. In the second, the math practice, participants were asked to solve a series of simple math operations (e.g., $5 * 2 + 1 = ?$). And in the third, participants performed the symbol recall while also solving the math problems, as they would do later in the actual testing phase. After the practice sections, participants were presented with the real trials, which consisted of a list of 15 sets of 3–7 randomized symbols that appeared intermixed with the equations, totaling 75 symbols and 75 math problems. At the end of each set, participants were asked to recall the symbols in the sequence they had been shown.

Following Unsworth et al. (2005), the total correct score reported by web-based test was used as the OSpan score. This score is the total number of correct symbols that were recalled in the correct order. The highest possible score was 75. Notice that the test also reported an absolute score, which refers to the sum of all perfectly recalled sets of symbols. However, the total correct score was used because the two scoring procedures have been found to correlate (Conway et al., 2005). Indeed, the two scores were highly correlated in the present work, $r(100) = .86$, $p = < .001$. In order to ensure that the OSpan task was performed adequately, a criterion of 80% accuracy (i.e., a maximum of 15 errors out of the 75 operations) on the math operations was set for all participants, as suggested by Turner and Engle (1989). The average accuracy was 93.9% (SD = 4.49%, range = 81%–100%). Therefore, all participants' data were included in the analyses.

5.3.2.2 Declarative memory

Participants' declarative memory capacity was assessed using two recall and recognition web-based tasks that indexed both verbal and nonverbal learning. The Modern Language Aptitude Test, subtest 5, and the Continuous Visual Memory Task served as verbal and nonverbal measures of declarative capacity, respectively. These tasks are described below.

(A) *The Modern Language Aptitude Test, Part 5, Paired Associates (MLAT5).* A web-based version of the Modern Language Aptitude Test, Part 5, Paired Associates, (MLAT5; J. B. Carroll & Sapon, 1959) was administered as a verbal measure of declarative memory capacity. The MLAT5 required participants to memorize artificial (i.e., pseudo-Kurdish) words and their meanings in English. Participants first studied 24-word association pairs for two minutes and then completed a two-minute practice section. During the practice section, the list of foreign words and their English equivalents were made available for participants to refer back if they needed to. Finally, participants completed a timed multiple-choice test (four minutes), in which they were asked to select the English meaning of each of the 24 pseudo-Kurdish words. For each correct response, one point was awarded, yielding a total score of 24 points. In total, the test duration was 8 minutes.

(B) *The Continuous Visual Memory Test (CVMT)*. A web-based version of the Continuous Visual Memory Task (CVMT; Trahan & Larrabee, 1988) was used as an assessment of nonverbal declarative memory capacity in the current investigation. The CVMT uses a visual recognition paradigm to assess nonverbal declarative memory, and involves asking participant to first view a series of complex abstract designs on the screen and then to indicate whether the image they just saw was novel ("new") in the series, or they had seen the image before ("old"). Seven of the designs were "old" (target items), and 63 were "new" (distractors). The target items appeared seven times (49 trials), and the distractors only once (63 trials) in the course of the task. All items were presented in a random but fixed order, each one appearing for two seconds. After the two seconds, participants were instructed to respond to the "OLD or NEW?" prompt on the screen by pressing either the "N" key for "NEW", or the "O" key for "OLD" on the keyboard. Overall, the CVMT required 10 minutes to administer. For each participant, a *d'*(d-prime) score (Wickens, 2002) for CVMT was computed. The *d'* score was used to account for the possible participants' bias toward choosing "OLD" or "NEW". The score was calculated from the z-score (i.e., standard deviation value) for the number of old items that were wrongly reported as new (false alarms) minus the z-score for the number of new items that were correctly reported as new (hits).

5.3.2.3 English reading proficiency test: The DIALANG test

The reading subtest of the web-based diagnostic test DIALANG (Alderson, 2005; Alderson & Huhta, 2005) was administered as a measure of participants' reading ability in English. This test has been used as a screening test for learner proficiency in previous L2 research (e.g., Alderson, Nieminen, & Huhta, 2016; Winke, 2013). The test took about 30 minutes to respond. It consisted of 30 multiple-choice items and assessed three reading subskills: i) understanding the main idea, ii) making inferences, and iii) understanding specific details (Harding, Alderson, & Brunfaut, 2015).

Based on the Common European Framework of Reference for Languages (CEFR; Council of Europe, 2001), the DIALANG reading test reports six scores bands (A1, A2, B1, B2, C1, C2). However, the version of the test that was incorporated into the WERTi interface for this investigation reported a number between 1 and 6. This number 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.

5.3.2.4 General English proficiency test: The C-test

The English general proficiency test used in this study was a web-based modified version of a former 30-minute timed placement test given to entering students at the *Fachsprachenzentrum* (Language Learning Center) at the University of Tübingen. The test consisted of a set of five short texts with different degrees of difficulty that were converted into C-test (gap filling) format. C-tests have been found to be good predictors of general language proficiency (Dörnyei & Katona, 1992; Eckes & Grotjahn, 2006; Harsch & Hartig, 2016). 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 by the test-taker. For example, if a word had 8 letters, 4 letters would be shown, and the remaining 4 letters had to be submitted. If a word had 9 letters in total, 4 letters would be given; the rest (5 letters) would have to be input. The first and last sentences of the text were, however, left intact (a sample C-test is shown in Figure 5.2). Four tests contained 25 gaps, and one text contained 24, amounting to 124 gaps. Each gap filled correctly was awarded one point, resulting in a maximum of 124 points. The C-test reliability (Cronbach's alpha) was .930.

At the beginning of the test, participants were informed of the nature, duration, and procedure of the proficiency test. Likewise, they were advised that i) consulting any offline or online materials while taking the test was not allowed; ii) hyphenated words counted as two words;

iii) words with apostrophes counted as one word, and thus the apostrophe counted as a character or letter; and iv) pressing the space bar at the end of a word would be counted as an additional character or letter, rendering the gap incorrect; therefore, this was to be avoided. The detailed instructions are reproduced in the appendix. During the test, learners could navigate through the five texts by clicking on a number (i.e., a web link) in the list (1-5) shown at the top of the page. Before moving to a different text, if there were still any empty gaps left in the text, participants received a reminder (in red) that not all of the gaps had been filled out. The empty gaps became highlighted (in red squares), and a "CANCEL" button appeared next to a "SAVE & GO TO NEXT" button. If they clicked on the "CANCEL" button, the highlighting disappeared. A "SUBMIT ALL & FINISH" button was available throughout the test in case participants wished to finish the test before completing all words. Once the gaps in the entire set of texts had been filled in, participants were notified and a "SUBMIT" button was shown. After submission, the test reported the total score, as well as the total number of correctly and incorrectly completed words in the five texts. This score was then mapped into one of the CEFR levels of proficiency using the following scale: A (basic; 0-64), B (intermediate; 65-84), and C (advanced; 85 and above).

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.

Figure 5.2. Sample C-test.

5.3.3 Pre- and posttests

To establish learners' semantic knowledge of English phrasal verbs before and after the treatments, two types of tests were designed. Considering the importance of assessing both productive and receptive mastery of L2 vocabulary (Schmitt, 2010), a cloze-like task to measure productive knowledge and a multiple-choice test to measure receptive knowledge were developed, respectively. These test formats were based on a study by Schmitt and Redwood (2011). As such, both instruments included similar items in the form of a single sentence testing the same target phrasal verb. The cloze-like task required participants to produce the targeted linguistic form themselves, which is more cognitively demanding than recognizing target items in a receptive test (Groot, 2000; Schmitt & Redwood, 2011). This type of form recall test has been widely used and has proven to be a good measure of lexical knowledge (Read, 1997; Schmitt & Redwood, 2011). Moreover, it has also been employed as a gauge of productive ability in recent studies investigating L2 multi-word units such as phrasal verbs (e.g., Garnier & Schmitt, 2016; Schmitt & Redwood, 2011) and collocations (e.g., González Fernández & Schmitt, 2015). In this test, the first letter of the verb and the particle was dispensed. For example, in the productive item for *point out* presented in Figure 5.3, the letters "p" and "o" were displayed.

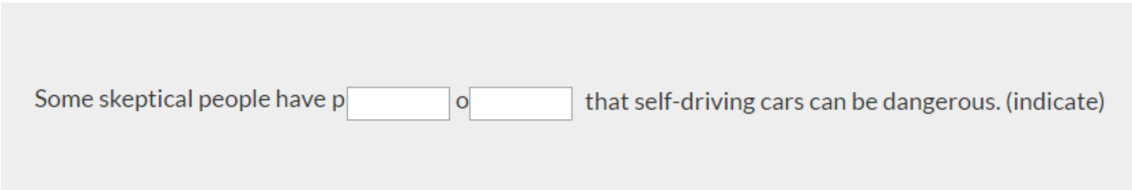


Figure 5.3. Productive item for *point out.*

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. There would be a possibility that participants remembered some of the multiple-choice answers in the receptive test had they taken the productive test second. It is feasible to assume that if the participants knew a phrasal verb productively, they would also know it receptively, as receptive knowledge is argued to precede productive knowledge (Daskalovska, 2011; Hopman & MacDonald, 2018; Laufer, 2012; Melka, 1997; Schmitt, 2010; Webb, 2008). A receptive version of the productive item for *point out* can be seen in Figure 5.4.

Each test contained 40 items, 30 assessed the target phrasal verbs (see appendix), and 10 verb-noun collocations (e.g., *break loose*) served as distractors. The sentence stimuli were primarily modeled after the sentence examples of phrasal verbs (e.g., *There is a debate going*

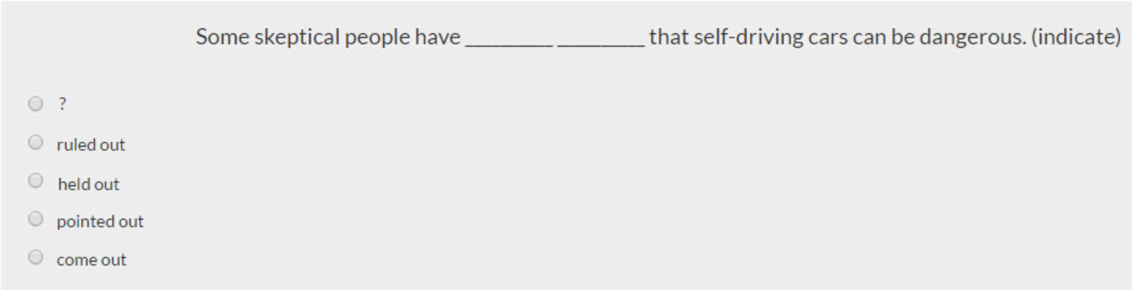


Figure 5.4. Receptive item for *point out.*

on right now between the two parties) from the PHaVE List (Garnier & Schmitt, 2015). The Internet as well as online and offline dictionaries were also consulted in order to generate the test sentences. The clarity and validity of the items were checked by two native speakers and two L2 experts. Based on the reviews, some of items were modified. In addition, both the receptive and productive tests were piloted with a group of 19 nonnative speakers. Minor alterations were made to the items after piloting. In order to avoid the potential threat to the internal validity of the construct measures if the pretest and the posttest were different (see Creswell, 2015), the same set of items was used for both tests, although they were presented in different random orderings, including the choices on the receptive test. Moreover, an expert on psychometrics and instrument development advised against using different items in the pretest and posttest, given that it was lexical knowledge what was to be assessed. Specifically, the individual meanings of phrasal verbs (i.e., 30 meanings) were to be tested, and thus the knowledge of these lexical items is specific rather than an abstract, generalizable ability. In this regard, Schmitt (2010) explains that

> [w]hen the construct being measured is a single rule or ability (third person-s, reading speed), it is possible to create tests where all the items address that single construct. However, vocabulary is largely item-based learning, and so each item addresses a separate construct, i.e. knowledge of a single lexical item (p. 185.)

In light of this, Schmitt argues that it is problematic to apply reliability methods such as internal consistency measures or item-total correlations to vocabulary tests. Consistent with the idea that vocabulary learning is essentially *item learning* (see also Sonbul & Schmitt, 2013), previous studies on L2 lexical development have also used the same set of items for the targeted vocabulary in both the pretest and the posttest, as in the current study (e.g., Alsadoon & Heift, 2015; S. Choi, 2017; Khezrlou et al., 2017).

Regarding the duration the tests, the productive test took a mean of 16.46 (*SD* = 10.46) minutes to complete in the pretest, and 13.57 (*SD* = 8.55) in the posttest. The mean duration of the receptive test was 6.40 (*SD* = 2.69) minutes in the pretest, and 5.44 (*SD* = 2.27) in the posttest.

Given the web-based nature of the experiment, participants were provided with a number of instructions related to the web-based testing environment, such as taking the tests in a quite place without interruption, completing the tests from start to finish in one sitting, avoiding refreshing or closing the browser page, or using the browser's back button. Importantly, they were asked not to take notes during the tests. The precise instructions are reproduced in the appendix. At the end of each test, learners were asked whether they had followed the instructions given at the beginning of the test, and whether they would recommend including their data in the analysis. They reported that instructions in the productive test were mostly followed in both the pretest (96%), and the posttest (98%). In the receptive test, participants indicated that instructions were always followed in the pretest (100%), and almost always in the posttest (99%). As for the inclusion of data in the analysis, 74% of participants recommended using their data from the productive test in the pretest, and 79% in the posttest. In the receptive test, 90% of participants recommended including their data in the pretest, and 88% in the pretest. However, all data were included in the analysis because the percentages for data inclusion were high overall. Moreover, this would also avoid losing any data.

Scoring of the receptive test was carried out automatically by the computer, as one point was awarded if the participant chose the correct option. The productive test, on the other hand, was manually scored for meaning by the researcher. 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. Likewise, the tense of the verb was not taken into account.

5.4 Procedure

Participants were predominantly recruited through the University of Tübingen mailing list. Other means of recruitment included online advertisement on Facebook and flyers placed in bulletin boards at the university. The recruitment message contained a link, that when clicked, took participants to the WERTi interface.

The web-based experiment consisted of i) a pretest block, ii) a cognitive testing block, iii) an over-two-week treatment period, and iv) a posttest block. On day 1, participants performed the pretest block. They were first shown a consent form in English. Following the ethics committee's recommendations, a consent form in German was also available through a link on the website. After giving informed consent, participants were taken with the registration window seen in Figure 5.5. At registration, a username, an email address, and a password were required. Participants were explicitly advised to choose a username that would not betray their real name or identity. Upon registering, a confirmation email with a link was automatically sent to the email address entered by the participants. After clicking on the confirmation link, learners were forwarded to the background questionnaire in the WERTi interface. The questionnaire recorded information about age, gender, educational level, and language background. It also requested information about the time participants spent reading on a computer, and the type of materials they most frequently read on this modality (see appendix). Once participants submitted the answers to the background questionnaire, they were first routed to the DIALANG reading test and subsequently to the C-test. Finally, participants took the productive and the receptive tests, in this order. To progress through these tests, a "NEXT" button was always present. Going back to the previous item was not possible. At the end of the receptive test, participants were asked to return to the interface the following day to continue with the experiment.

On day 2, participants completed the cognitive testing block. In this block, the three cognitive tests under investigation were administered in a fixed order: OSpan, CVMT, and MLAT5. The instructions given to the participants at the beginning of the block are reproduced in the appendix. After submitting their answers to the last test (MLAT5), participants were instructed to resume the experiment on the next day. From the third day onwards and for over two weeks, all learners were instructed to read at least 2 (up to 5) texts per day via the WERTi interface, and to complete a short questionnaire after each test. During the over-two-week treatment period, learners read 40 news articles, and their interactions with the system were recorded.

Upon finishing reading the 40 articles, participants were directed to the posttest block, in which they completed the productive and receptive tests for the second time, and filled in a debriefing questionnaire (see description below). After submitting the answers to the debriefing

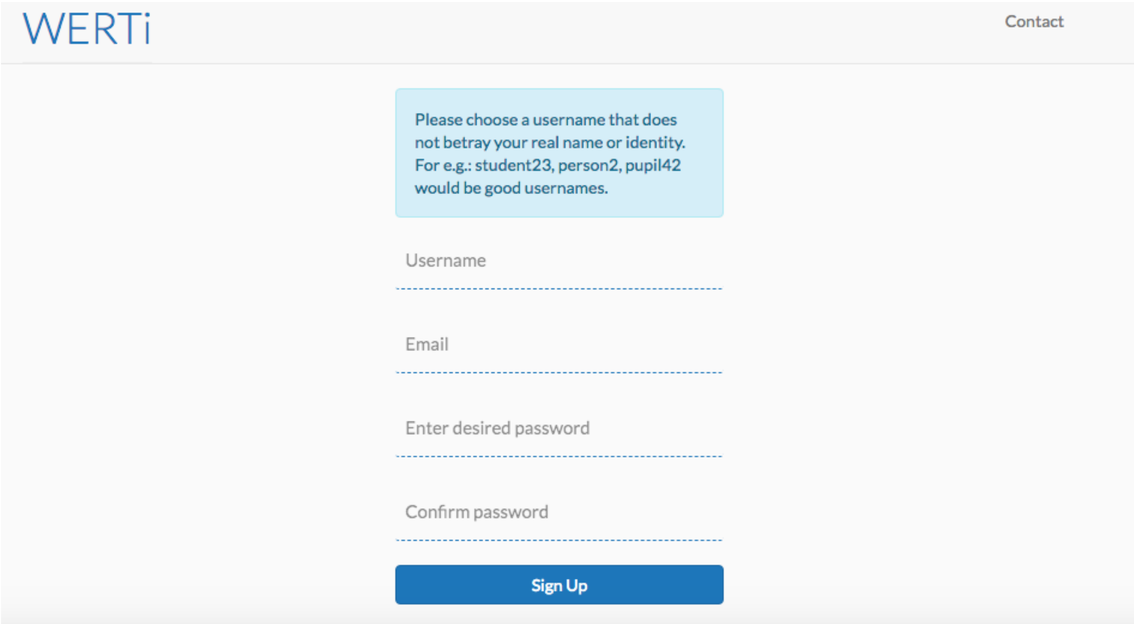


Figure 5.5. Registration window.

questionnaire, compensation was sent via email in the form of an electronic gift card code. About two weeks after the commencement of the study and throughout the following week, reminders were sent to those participants who had not finished the experiment within the allotted amount of time.

5.4.1 Treatments

All participants were asked to read the pre-selected 40 authentic news articles taken from the Reuters website by logging onto the WERTi interface using their username and password. Once participants logged in, they were presented a webpage with instructions according the treatment condition, and a hyperlink with the title of the news article to be read. On this page, both groups were instructed to read the article carefully. Nonetheless, learners in the form-focused group were further requested to complete the blank spaces in the text by clicking on the boxes, and selecting the correct option from a pulldown menu. For this group, after clicking on the news title, the system returned an enhanced webpage in which the particles of English phrasal verbs (e.g., *out*, *off*, *in*) had been automatically turned into multiple-choice options (Figure 5.7). For the meaning-focused group, the system returned a webpage without any enhancement (Figure 5.6), except for a clickable box located at the end of the news pages to allow the learners to render the reading questionnaire visible after reading the text. The precise instructions are reproduced in the appendix.

As illustrated in Figure 5.7, once the empty box was clicked, a dropdown menu with four options was displayed. One option was the correct answer, and the others were distractors. Two of the distractors were retrieved from a list of semantically similar words automatically gathered using distributional semantics methods (i.e., automatic vocabulary extraction based on co-occurrence information; Skeppstedt, Ahltorp, & Henriksson, 2013). A third distractor was fetched from a list of manually collected words that function similarly to particles (e.g., prepositions). No distractors were repeated in the dropdown menus. If the answer was correct, it turned green. If the answer was incorrect, it was highlighted in red. The box also became red and an “X” (also in red) appeared. 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. Participants were instructed to only use the smiley face as a last resort. At the bottom of each article page, a button to continue to a reading questionnaire after having read the article was displayed for both groups. In the case of the form-focused group, however, the button only became activated after all of the blanks in the text had been filled out.

Figure 5.6. Meaning-focused condition.

Figure 5.7. Form-focused condition.

5.4.2 Reading questionnaire

The reading questionnaire had two parts, one assessing global understanding of the text, and the other collecting participants' feedback.

Part I. *Global reading comprehension.* The first part consisted of a multiple-choice question in which learners were required the select the main (top-level) idea of the article they had just read (see sample question shown in Figure 5.8). The distractor options represented minor details in the article (i.e., low-level ideas; see Carrell, 1992).

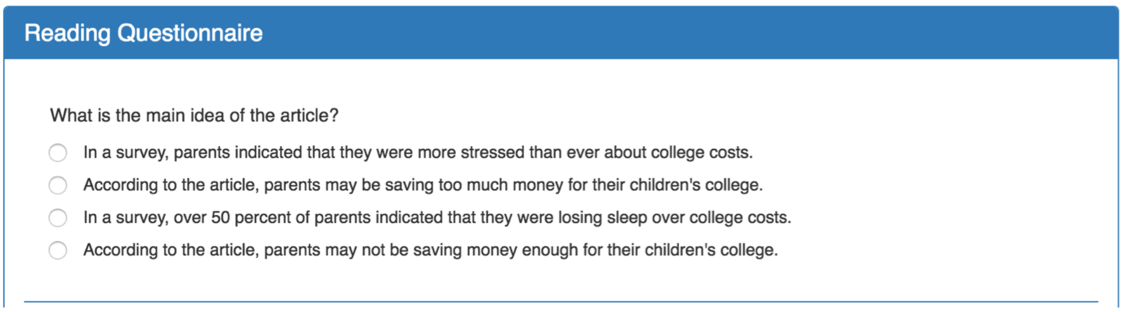


Figure 5.8. Sample main idea question.

Part II. *Feedback questionnaire.* As can be seen in Figure 5.9, in the second part of the questionnaire, participants reported perceived text difficulty (ranging from 1 = *very easy* to 5 =

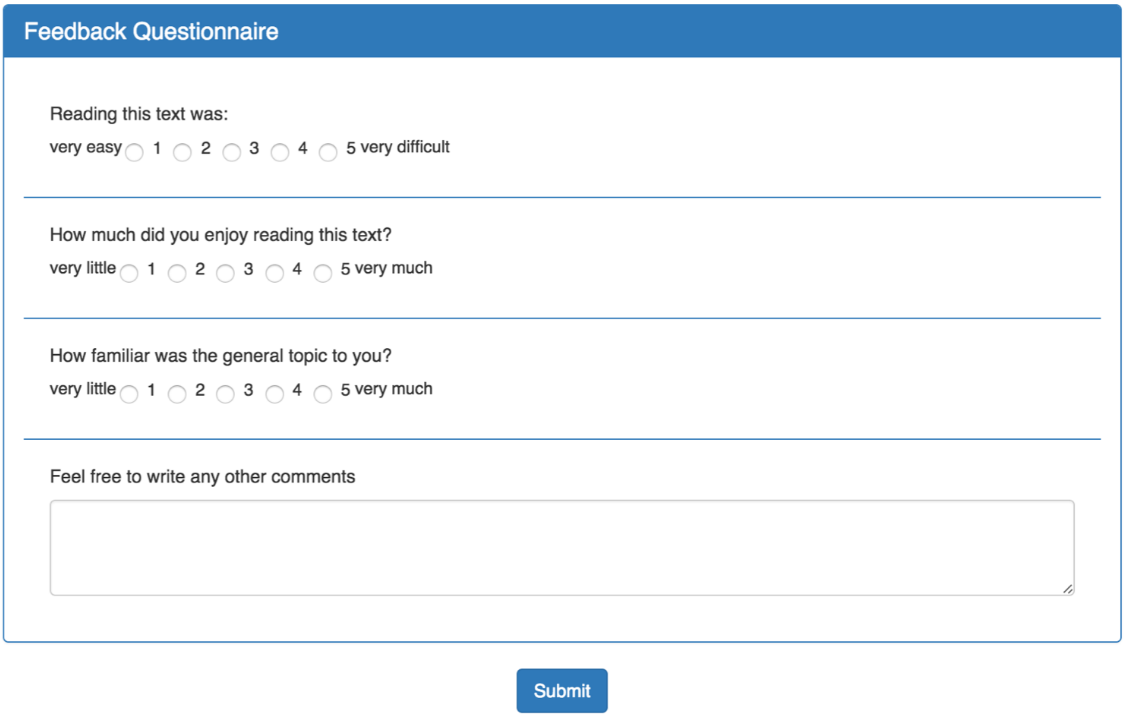


Figure 5.9. Feedback questionnaire.

very difficult), enjoyment of the text, and familiarity with text topic (ranging from 1 = *very much* to 5 = *very little*; Pino-Silva, 1992, 2009). This part allowed to ensure that the texts were relatively comparable across groups. There was also space for comments and a "Submit" button. The article was always visible while participants completed the questionnaire.

5.4.3 Debriefing questionnaire

The debriefing questionnaire included two parts, one part contained questions concerning the study, and the other contained questions concerning the technology background of the participants. In the first part, all participants were asked about what they thought the aim of the study was. Using a scale ranging from 1 (*Never*) to 5 (*Always*), learners in the form-focused group were further questioned about whether they had noticed the visual enhancement of the English phrasal verbs when reading the texts. Likewise, on a scale ranging 1 (*Not at all*) to 5 (*Very much*), they indicated whether the enhancement had helped them pay attention to the use of English phrasal verbs while reading, and whether they had felt any improvement in using English phrasal verbs after reading the 40 texts. Finally, learners were asked how the enhancement impacted their reading of the 40 texts, whether it made the reading easier or more difficult, or whether it had any impact at all. In the second part, both groups were asked whether they owned a computer; how long they had been using a computer; how many hours they used a computer in a typical day; where they used a computer; and whether they felt comfortable using a computer. They were also asked whether they had participated in non-traditional learning formats, and if they had, how much learning experience they had; and whether they had taken any web-based courses. Lastly, participants were questioned on what they liked and disliked about the experience of reading texts on the web during the study. The detailed instructions are reproduced in the appendix.

5.4.4 Overview of procedure

An overview of the procedure is presented in Figure 5.10.

5.5 Data analysis

Data were analyzed by descriptive statistics, *t* tests, correlations, and linear mixed-effects logistic regression modeling using the statistical software package R version 3.3.2 (R Development Core Team, 2016). To construct logistic mixed-effects models, the glmer (generalized linear mixed model) function in the *lme4* package (D. Bates, Mächler, et al., 2015) was employed. In order to avoid nonconvergence, the BOBYQA algorithm was used as the optimizer, as recommended by Linck and Cunnings (2015).

Mixed-effects regression modeling does not require prior averaging, and allows fitting both fixed and random effects at the same time (Baayen, 2008; Baayen, Davidson, & Bates, 2008; Jaeger, 2008). A fixed effect is a difference between observations with different values on an independent variable, with different levels that provide more information than just being different identities (e.g., treatment vs. control groups). A random effect, on the other hand, is the variance of effects that depends, in principle, on randomly selected observations (e.g., participants, words). In other words, fixed effects account for variation resulting from experimental factors, whereas random effects account for variation resulting from different observations that cannot be

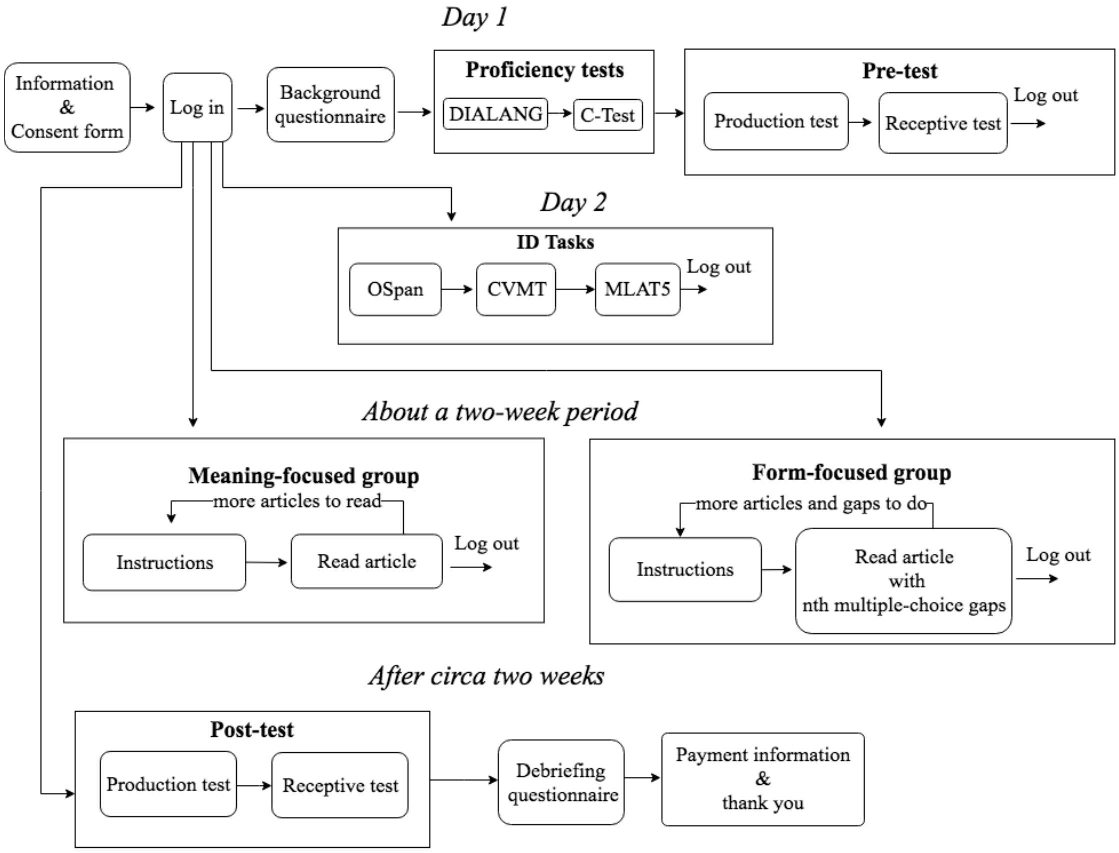


Figure 5.10. Overview of procedure.

distinguished beyond their being different observations.

The parameter estimates or regression coefficients of the logistic mixed-effects models are expressed in log odds probability units, or logits, that is, natural logs of odds ratios of a correct response. Logits are comparable to proportions but satisfy the mathematical prerequisites of linear models (Stolarova et al., 2016). In this dissertation, an odd ratio was the ratio of the probability of obtaining a 1 (correct response) to the probability of obtaining a 0 (incorrect response). Hence, positive (higher logits) estimates indicated a greater likelihood of accuracy while negative estimates (lower logits) represented a lesser likelihood of accuracy. The intercept is the log odds ratio at the point where all the other variables are zero. The effects are the estimates of the differences in performance between the values (i.e., levels) captured in a particular variable for those cases which have zeros in all other variables (Cohen, Cohen, West, & Aiken, 2003). For instance, the effect of the variable *group* is the effect of *group* if all the other variables have the value of zero.

If there is no interaction, this would be also the effect of *group* across any of the other variables (see below).

The use of mixed-effects regression modeling as a tool for statistical analysis of L2 data has been increasingly recommended (Cunnings, 2012; Cunnings & Finlayson, 2015; Linck & Cunnings, 2015; Murakami, 2016; Plonsky, 2017; Plonsky & Oswald, 2017). Notably, mixed-effects models are particularly well suited for the analysis of individual differences that affect L2 outcomes, as they offer the flexibility to model the interactive and main effects of multiple factors, while also considering the inherent variability of learners and materials (Linck, 2016). Mixed effects are also robust with respect to missing data (Baayen, 2008; Jaeger, 2008). In the present dissertation, mixed-effects models were used to examine how individual differences, namely working memory and declarative memory, influenced outcomes in a web-based L2 vocabulary learning study supported by NLP. Mixed-effects modeling has also been suggested for the analysis of L2 longitudinal data (i.e., data gathered from the same individual on several occasions; Cunnings & Finlayson, 2015). Consequently, mixed-effects models were also applied to estimate the development of accuracy of the form-focused group during the over-two-week treatment period. In what follows, a description of the data analysis performed on the accuracy data collected from both the receptive and productive tests as well as the treatment period is given.

5.5.1 Receptive and productive test data

Given the binomial distribution of the dependent variable (i.e., each response has a fixed probability of being correct), logistic mixed-effects models were used to predict accuracy (correct vs. incorrect) in both receptive and productive tests, as these models are recommended to account for this type of nonnormal distribution (cf. Jaeger, 2008; Quené & van den Bergh, 2008). Specifically, the dependent variable was a binary response to denote whether each item was correct (success; 1 point), or incorrect (failure; 0 points). The scores from the cognitive tests (OSpan, MLAT5, and CVMT), the learning condition (*group* factor), the type of knowledge assessed (*knowledge type* factor), and the pretest and posttest measures (*test type* factor) were considered as fixed effects, and participants and phrasal verbs were treated as random crossed effects, i.e., random intercepts for participants and phrasal verbs (Baayen et al., 2008; Cunnings, 2012). The *group* factor included two levels (meaning-focused vs. form-focused); the *knowledge type* factor contained two levels (receptive vs. productive); and the *test type* factor had two levels (pretest vs. posttest). In accordance with pertinent literature (Aiken & West, 1991), the raw scores of the continuous variables *working memory* (OSpan) and *verbal declarative memory* (MLAT5) were mean-centered, or centered, i.e., the mean of each score was subtracted from each individual's score, yielding means that equaled zero. Centering the numerical variables was used to minimize multicollinearity (high correlations among predictors) that may occur when computing interaction terms, as well as to simplify interpretability of analyses (Aiken & West, 1991). As noted earlier, the CVMT *d'* scores were calculated from z-scores, meaning that scores already had a mean of zero. Therefore, no transformation of the continuous variable *nonverbal declarative memory* was required for the analysis.

The procedure to conduct the mixed-effects analysis was as follows. Firstly, all factorial predictors were recoded in order to facilitate the interpretation of the results and to additionally reduce multicollinearity (Linck & Cunnings, 2015, p. 192). This was carried out using contrast coding, that is, the values –.5 (absence) and .5 (presence) were assigned to the levels of each variable, so that the zero value is the mean between these two values (see Cohen et al., 2003, for detailed description of coding systems). The contrast-coded fixed effects were as follows: *type of knowledge* (–.5 = productive, .5 = receptive), *type of test* (–.5 = posttest, .5 = pretest); and *group* (–.5 = meaning-focused, .5 = form-focused). In the models, the intercept represented the log odds for the combination of the levels coded as –.5. Therefore, the effect of a given factor (i.e., the difference in performance) was estimated from the comparison between the level of the factor coded as –.5 and the level of the factor coded as .5 (Linck, 2016).

Secondly, regression models containing higher-order interaction terms were built. In order to achieve a parsimonious analysis (D. Bates, Kliegl, Vasishth, & Baayen, 2015; Matuschek, Kliegl, Vasishth, Baayen, & Bates, 2017), only one cognitive predictor was entered into each model. Higher-order interactions were included to determine whether there were any associations of the predictors with performance. Higher-order interactions involve more than two variables, allowing to examine whether the association between two variables is a function of the different levels of one or more other variables (Elliott, 1998). In other words, if there is an interaction, the estimates, the regression coefficients for the individual predictors, are conditional effects (i.e., simple slopes): separate effects of one variable within levels of another variable (Cohen et al., 2003). In the present investigation, if an interaction was found to be significant, it was resolved by further carrying out a simple slope analysis, as recommended by Aiken and West (1991). A simple slope analysis entails resolving the interaction and differentiating between the effect of some variable at different values of another variable. In order to resolve or unpack the interaction, a recoding procedure was applied (Cohen et al., 2003), thereby running two separate multiple regression analyses with the same data, only altering the coding (0 and 1) of the variables involved in the interaction. In the case of an interaction involving a continuous variable, one standard deviation (*SD*) was added to and subtracted from the average mean of the continuous variable. This would enable to further assess the simple effect at the levels of both low (1 *SD* below the mean) and high (1 *SD* above the mean) performance. If there were no significant interactions, there was no need to differentiate between the variables, signifying that the effect of a given predictor was statistically the same across the other predictors. This average effect is considered a main effect, that is, an effect that is not limited by any other predictors and thus applies across all data. The flowchart in Figure 5.11 shows the procedure followed to perform mixed-effects data analysis.

5.5.2 Treatment data

Two mixed-effects analyses on the accuracy data from the treatment period were undertaken. One analysis was done on the data derived from the main idea question in the reading questionnaire. This analysis aimed to determine the combined influence of working memory, treatment condition, and English reading proficiency on performance. It also sought to examine whether the enhancement of the phrasal verbs via multiple-choice tasks had any detrimental effect on reading comprehension. As before, accuracy was used as a binary outcome (correct = 1, incorrect = 0). The levels of reading proficiency reported by the DIALANG test (see section 1.3.3.1) made up the *English reading proficiency* factor, and as such it contained two levels: advanced vs. intermediate. The *English reading proficiency* factor, the *group* factor, and the centered OSpan score were modeled as fixed effects, and participants and items were fitted as crossed random effects. The

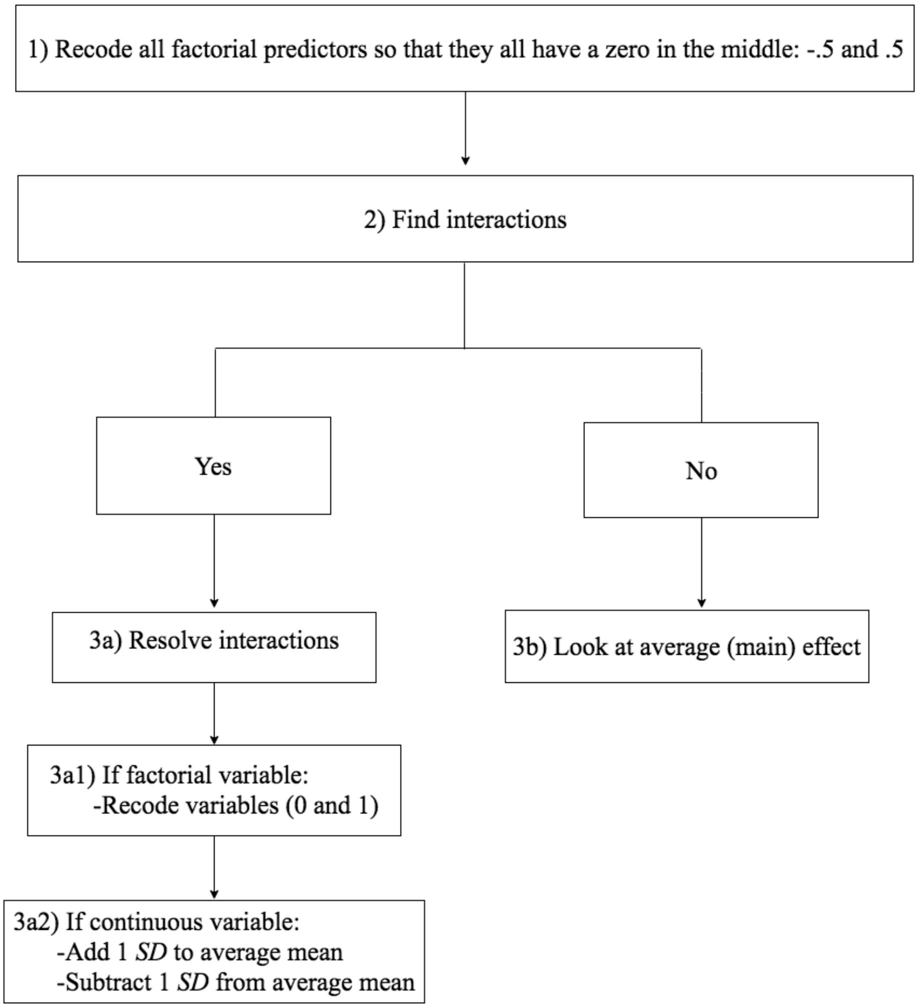


Figure 5.11. Procedure of the mixed-effects data analysis.

procedure to complete the analysis on the main idea accuracy data was the same as for the receptive and productive data: i) the factorial predictor *English reading proficiency* was contrast-coded (–.5 = advanced, .5 = intermediate); and ii) higher-order interactions were examined.

Another mixed-effects analysis was conducted to model accuracy in the multiple-choice activity in the form-focused group. More specifically, the analysis was carried out to assess whether time as a variable predicted performance. Since the 40 texts were presented in a fixed order, the text sequence (1-40) was used as a proxy to represent the variable *time*. In the logistic mixed-effects models, accuracy was expressed as binomial: correct responses = successes, and incorrect responses and cheat instances = failures. To account for possible nonlinearity in L2 development (Lightbown, 1983; Murakami, 2016), *time* as both a linear and a quadratic (i.e., parabola) predictor was included using orthogonal polynomial coding (cf. Hutcheson, 2011).

5.6 Summary

This chapter described the methods used to investigate the research questions. Moreover, the procedures for data collection and statistical analyses were explained. The next chapter presents the results of these analyses.

Chapter 6

Results

This chapter reports on the results of the study. It first gives the descriptive statistics and *t*-test results for the percentage scores of both receptive and productive tests. Next both descriptive and *t*-test statistics for the cognitive tests under investigation, as well as a correlational analysis among these cognitive variables are provided. This is followed by the combined mixed-effects analyses of the cognitive measure data and the receptive and productive test accuracy data. Then, the results of the computer log data analyses are addressed. The chapter concludes with a report of the debriefing questionnaire results.

6.1 Learning effects: Pretest vs. posttest scores

6.1.1 Receptive test

Table 6.1 presents the descriptive statistics summarizing participants' performance on the receptive test. In order to determine whether the meaning-focused and the form-focused groups had differences in prior receptive knowledge of the targeted phrasal verbs at the start of the study, an independent-samples *t* test was performed on pretest scores. Results showed that there was no statistical difference between the groups, $t(124.94) = 0.533$, $p = .595$, $d = .14$.

In order to test whether the treatment condition resulted in absolute learning gains (see

Plonsky, 2017), pretest–posttest contrasts were carried out. Results from paired-samples *t* tests showed that there was a statistical difference between pre and post testing on the receptive test for both the meaning-focused group, $t(62) = -3.602$, $p < .001$, $d = -.45$, and form-focused group, $t(63) = -4.287$, $p < .001$, $d = -.54$. Further analysis with independent samples *t* tests revealed that the difference in posttest scores between the groups was nonsignificant, $t(122.4) = 0.817$, $p = .416$, $d = .14$.

Table 6.1

Descriptive statistics for pretest and posttest comparison of percentage scores on the receptive test.

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

6.1.2 Productive test

Table 6.2 provides the descriptive statistics of participants' performance on the productive test. In order to determine whether the meaning-focused and the form-focused groups had differences in prior productive knowledge of the targeted phrasal verbs at the commencement of the study, an independent-samples *t* test was performed on pretest scores. Results showed that there was no statistical difference between the groups, $t(121.1) = 1.229$, $p = .222$, $d = .22$.

In order to test whether the treatment condition resulted in absolute learning gains, pretest–posttest contrasts were carried out. Results from paired-samples *t* tests showed that there was a statistical difference between pre and post testing on the productive test for both the meaning-focused group, $t(62) = -7.485$, $p < .001$, $d = -.94$, and the form-focused group, $t(63) = -8.314$, $p < .001$, $d = -1.04$. Further analysis with independent samples *t* tests revealed that the difference in postest scores between the groups was nonsignificant, $t(124.96) = 1.620$, $p = .108$, $d = .29$.

Table 6.2

Descriptive statistics for pretest and posttest comparison of percentage scores on the productive test.

Test		Meaning-focused		Form-focused	
		Pre	Post	Pre	Post
Productive	*M*	44.09	54.73	47.07	59.12
	SD	15.85	15.66	13.44	16.18

6.2 Individual difference measures

6.2.1 Descriptive statistics

Descriptive statistics for the three cognitive tests under investigation (OSpan, MLAT5, CVMT) are reported below.

6.2.1.1 Working memory

Descriptive statistics for the OSpan are reported for all participants and both instructional conditions (see Table 6.3). An independent-samples *t* test indicated that scores were not significantly different between the meaning-focused and the form-focused groups, $t(74.345) = 1.184$, $p = .240$, $d = .25$.

Table 6.3

Descriptive statistics for OSpan scores.

Test		All (N = 102)	Meaning-focused (n = 41)	Form-focused (n = 61)
OSpan	*M*	22.95	25.05	21.54
	SD	14.15	15.73	12.96

6.2.1.2 Verbal declarative memory

Descriptive statistics for MLAT5 are reported for all participants and both instructional conditions (see Table 6.4). An independent-samples *t* test indicated that scores were significantly different between the meaning-focused and the form-focused groups, $t(89.387) = 2.334$, $p = .022$, $d = .47$.

Table 6.4

Descriptive statistics for MLAT5 scores.

Test		All (N = 92)	Meaning-focused (n = 40)	Form-focused (n = 52)
MLAT5	*M*	17.15	18.40	16.19
	SD	4.80	3.75	5.32

6.2.1.3 Nonverbal declarative memory

Descriptive statistics for CVMT (d-prime scores) are reported for all participants and both instructional conditions (see Table 6.5). An independent-samples *t* test indicated that scores were not significantly different between the meaning-focused and the form-focused groups, $t(84.399) = 0.726$, $p = .470$, $d = .15$.

Table 6.5

Descriptive statistics for CVMT d' scores.

Test		All (N = 89)	Meaning-focused (n = 38)	Form-focused (n = 51)
CVMT	*M*	1.96	2.01	1.93
	SD	0.53	0.49	0.56

6.2.2 Correlations

Considering both groups, Pearson's correlations to assess relationships between the cognitive measures were performed (see Table 6.6). The analysis revealed that raw scores of the OSpan were significantly correlated to both the raw scores of the MLAT5, $r(90) = .23$, $p = .028$, and the CVMT *d'* scores, $r(87) = .37$, $p < .001$, respectively. The raw scores of the MLAT5 and the CVMT *d'* scores were not significantly correlated to each other, $r(84) = .13$, $p = .234$). The significant correlations found were not of concern for multicollinearity because only one cognitive measure at a time was included as a predictor in each model (see Chapter 5).

Table 6.6

Correlations for cognitive measures.

Test	OSpan	MLAT5
OSpan	–	
MLAT5	.23	–
CVMT	.37	.13

Note. OSpan = Automated Operation Span Task; MLAT5 = Modern Language Aptitude Test, Part 5; CVMT = Continuous Visual Memory Task.

6.2.3 Regression analysis

Mixed-effects models were used to estimate the effects of working and declarative memory on learning English phrasal verbs both receptively and productively. The results of the mixed-effects analysis are presented below.

6.2.3.1 Working memory

As shown in Table 6.7, there was a significant three-way interaction between group, test type, and participants' individual differences in working memory capacity (estimate = 0.023, *SE* = 0.008, $p = .004$), suggesting that the effect of working memory on learning depended on the group. 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, $p = .024$). In line with the t-test analysis, performance was statistically equivalent across groups (estimate = 0.246, *SE* = 0.202, $p = .224$).

As for random effects, phrasal verbs explained less variance than participants, given that their random intercept did not vary as much around the overall intercept (*SD* = 0.951). Participants, on the other hand, showed more variance around the intercept (*SD* = 1.605), denoting more variability in performance, as explained by individual differences between participants.

Table 6.7

Logistic mixed-effects regression with working memory (OSpan) as an individual difference predictor.

Fixed effect	Estimate	*SE*	*p*
Intercept	1.496	0.311	<.001
Knowledge type	2.757	0.066	<.001
Working memory	0.010	0.007	.164
Test type	0.659	0.058	<.001
Group	0.246	0.202	.224
Knowledge type:Working memory	-0.001	0.004	.844
Knowledge type:Test type	-0.231	0.115	.045
Working memory:Test type	0.002	0.004	.687
Knowledge type:Group	-0.093	0.118	.428
Working memory:Group	0.033	0.014	.019
Test type:Group	0.116	0.115	.314
Knowledge type:Working memory:Test type	0.003	0.008	.727
Knowledge type:Working memory:Group	0.025	0.008	.002
Knowledge type:Test type:Group	0.166	0.230	.470
Working memory:Test type:Group	0.023	0.008	.004
Knowledge type:Working memory:Test type:Group	0.005	0.016	.759
Random effects	Variance	*SD*	
Participant	0.905	0.951	
Phrasal verb	2.576	1.605	

Note. All factors were contrast-coded, as follows: Type of knowledge (–.5 = productive, .5 = receptive), Type of test (–.5 = pretest, .5 = posttest), Group (–.5 = meaning-focused, .5 = form-focused).

6.2.3.2 Verbal declarative memory

As seen in Table 6.8, there was no significant three-way interaction between group, test type, and participants' individual differences in verbal declarative memory capacity (estimate = 0.018, *SE* = 0.028, *p* = .522). Results also revealed that there was no significant evidence for an effect of verbal declarative memory on learning (estimate = -0.005, *SE* = 0.014, *p* = .705).

Table 6.8

Logistic mixed-effects regression with verbal declarative memory (MLAT5) as an individual difference predictor.

Fixed effect	Estimate	*SE*	*p*
Intercept	1.579	0.319	<.001
Knowledge type	2.856	0.073	<.001
Verbal declarative memory	0.038	0.025	.125
Test type	0.716	0.064	<.001
Group	0.411	0.215	.056
Knowledge type:Verbal declarative memory	-0.026	0.015	.075
Knowledge type:Test type	-0.259	0.126	.040
Verbal declarative memory:Test type	-0.005	0.014	.705
Knowledge type:Group	-0.011	0.130	.932
Verbal declarative memory:Group	0.069	0.049	.160
Test type:Group	0.220	0.126	.081
Knowledge type:Verbal declarative memory:Test type	0.006	0.028	.824
Knowledge type:Verbal declarative memory:Group	0.076	0.029	.009
Knowledge type:Test type:Group	0.244	0.252	.334
Verbal declarative memory:Test type:Group	0.018	0.028	.522
Knowledge type:Verbal declarative memory:Test type:Group	-0.056	0.056	.317
Random effects	Variance	*SD*	
Participant	0.882	0.939	
Phrasal verb	2.695	1.642	

Note. All factors were contrast-coded, as follows: Type of knowledge (–.5 = productive, .5 = receptive), Type of test (–.5 pretest, .5 = posttest), Group (–.5 = meaning-focused, .5 = form-focused).

6.2.3.3 Nonverbal declarative memory

As observed in Table 6.9, there was a significant three-way interaction between group, test type, and and participants' individual differences in nonverbal declarative memory (estimate = 0.456, *SE* = 0.232, *p* = .050), suggesting that the effect of nonverbal memory on learning depended on the group. 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).

Table 6.9

Logistic mixed-effects regression with nonverbal declarative memory (CVMT) as an individual difference predictor.

Fixed effect	Estimate	*SE*	*p*
Intercept	1.034	0.532	.052
Knowledge type	2.874	0.242	<.001
Nonverbal declarative memory	0.279	0.213	.191
Test type	0.697	0.235	.003
Group	0.697	0.868	.422
Knowledge type:Nonverbal declarative memory	-0.025	0.118	.830
Knowledge type:Test type	-0.181	0.470	.7
Nonverbal declarative memory:Test type	0.031	0.116	.789
Knowledge type:Group	0.059	0.477	.901
Nonverbal declarative memory:Group	-0.219	0.426	.608
Test type:Group	-0.767	0.470	.103
Knowledge type:Nonverbal declarative memory:Test type	-0.015	0.232	.947
Knowledge type:Nonverbal declarative memory:Group	-0.039	0.236	.868
Knowledge type:Test type:Group	-0.447	0.937	.633
Nonverbal declarative memory:Test type:Group	0.456	0.232	.050
Knowledge type:Nonverbal declarative memory:Test type:Group	0.337	0.463	.467
Random effects	Variance	*SD*	
Participant	0.948	0.974	
Phrasal verb	2.807	1.675	

Note. All factors were contrast-coded, as follows: Type of knowledge (–.5 = productive, .5 = receptive), Type of test (–.5 = pretest, .5 = posttest), Group (–.5 = meaning-focused, .5 = form-focused).

6.3 Computer log data

In order to provide a fine-grained, incremental assessment of learner development during both meaning-focused and form-focused treatments, the learner log data were analyzed. The results of this analysis are given below.

6.3.1 Development of accuracy in the form-focused group

The scatterplot in Figure 6.1 gives an overall impression of the development of the accuracy with which individual learners completed the multiple-choice activities in the form-focused group (n = 64). Each mark corresponds to a document 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 (M = 132.26, SD = 86.78).

In order to determine whether *time* as a variable predicted performance in the multiple-choice gaps (see Table 6.10), a mixed-effects analysis 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 the texts 20-21 (approximately half the treatment time) and then declined. To further investigate whether this effect was still attested after removing learners who steadily scored at ceiling, a mixed-effects analysis with only those consistently scoring below 95% (n = 20) was carried out. For the latter learners, as seen in Table 6.11, there was a significant linear decrease in accuracy over time (estimate = -0.4766, SE = 0.204, p = .019).

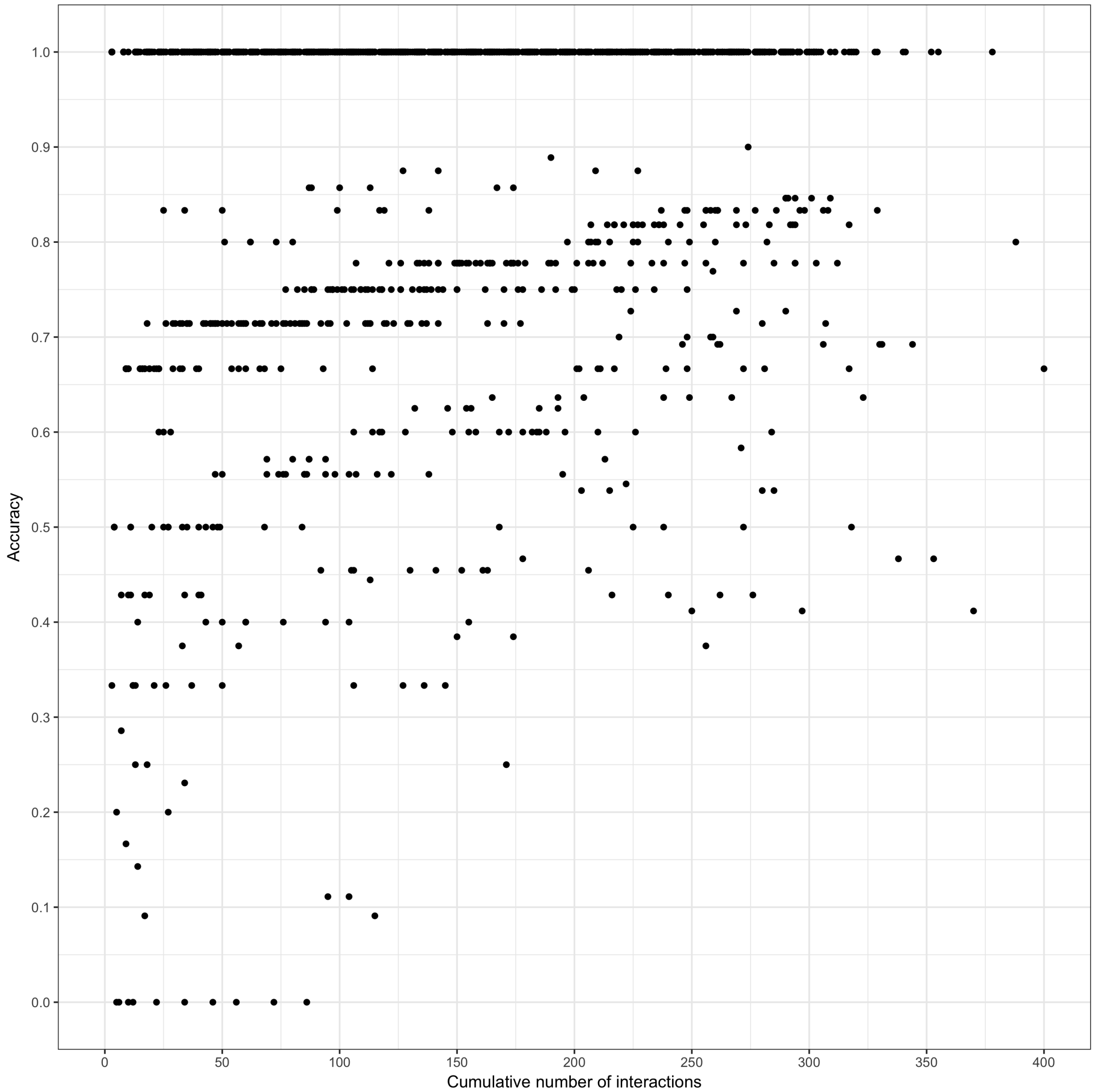


Figure 6.1. Scatterplot of accuracy by interacted items.

Table 6.10

Logistic mixed-effects regression with time as a predictor of accuracy in the multiple-choice activities (n = 64).

Fixed effect	Estimate	*SE*	*p*
Intercept	4.111	0.311	<.001
Time linear	0.015	0.187	.938
Time quadratic	-0.385	0.184	.037
Random effects	Variance	*SD*	
Participant	5.4	2.324	

Table 6.11

Logistic mixed-effects regression with time as a predictor of accuracy in the multiple-choice activities (n = 24).

Fixed effect	Estimate	*SE*	*p*
Intercept	1.495	0.180	<.001
Time linear	-0.477	0.204	.019
Time quadratic	-0.322	0.202	.110
Random effects	Variance	*SD*	
Participant	0.621	0.788	

6.3.2 Reading questionnaire

6.3.2.1 Part I. Global reading comprehension

To examine whether individual differences in cognitive abilities, reading proficiency in English, and instructional conditions had a combined effect on global reading comprehension (main ideas), mixed-effects models with three-way interaction terms involving these variables were developed (Elliott, 1998).

The results in Table 6.12 show that the three-way interaction of group, English reading proficiency, and working memory was not significant according to conventional standards (estimate = -0.034, *SE* = 0.019, p = .072). However, there was a significant interaction between the group factor and the English reading proficiency factor (estimate = -0.854, *SE* = 0.272, p = .002), signifying that the effect of the instructional condition captured by the group variable was different depending on participants' reading proficiency level. The results of a simple slope analysis (Aiken & 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). In addition, results showed that there was no evidence for a main effect of working memory on accuracy in the global reading comprehension measure (intercept = 0.006, *SE* = 0.005, *p* = .219).

Regarding random effects, participants explained less variance than items, given that their random intercept did not vary as much around the overall intercept (*SD* = 0.472). Items, on the other hand, showed more variance around the intercept (*SD* = 1.396), denoting more variability in performance, as explained by item-by-item differences.

Table 6.12

Logistic mixed-effects regression for reading comprehension with working memory (OSpan) as an individual difference predictor.

Fixed effect	Estimate	*SE*	*p*
Intercept	0.518	0.229	.023
Group	-0.189	0.136	.163
English reading proficiency	-0.558	0.136	<.001
Working memory	0.006	0.005	.219
Group:Reading proficiency	-0.854	0.272	.002
Group:Working memory	0.004	0.009	.646
English reading proficiency:Working memory	-0.008	0.009	.398
Group:English reading proficiency:Working memory	-0.034	0.019	.072
Random effects	Variance	*SD*	
Participant	0.223	0.472	
Item	1.948	1.396	

Note. All factors were contrast-coded, as follows: Group (–.5 = meaning-focused, .5 = form-focused), English reading proficiency (–.5 = advanced, .5 = intermediate).

As for both verbal declarative memory and nonverbal verbal declarative memory, further mixed-effects analyses indicated that the three-way interactions involving these variables were not significant (group by English reading proficiency by verbal declarative memory: estimate = 0.018, *SE* = 0.063, *p* = .768; group by English reading proficiency by nonverbal declarative memory: estimate = -0.421, *SE* = 0.515, *p* = .414). Additionally, there was no evidence for a main effect of either verbal declarative memory (estimate = 0.005, *SE* = 0.016, *p* = .747), nor nonverbal declarative memory (estimate = 0.129, *SE* = 0.129, *p* = .314) on accuracy in the global reading comprehension measure.

6.3.2.2 Part II. Feedback questionnaire

Descriptive statistics for the feedback questionnaire are reported for all participants and both instructional conditions (see Table 6.13). Independent-samples *t* tests indicated that there were no significant differences between groups in terms of text difficulty, $t(124.48) = 0.375$, $p = .709$, $d = .07$; enjoyment, $t(124.48) = 0.375$, $p = .709$, $d = .07$; or familiarity, $t(124.33) = 0.257$, $p = .798$, $d = .05$.

Table 6.13

Descriptive statistics for feedback questionnaire.

Category		All (*N* = 127)	Meaning-focused (*n* = 63)	Form-focused (*n* = 64)
Difficulty	*M*	2.93	2.91	2.94
	SD	1.23	1.24	1.23
Enjoyment	*M*	2.93	2.91	2.94
	SD	1.23	1.94	1.23
Familiarity	*M*	2.35	2.34	2.36
	SD	1.25	1.25	1.25

Note. Scale from 1 to 5; Difficulty: 1 = very easy, 5 = Very difficult; Enjoyment and Familiarity: 1 = Very little; 5 = Very much.

6.4 Debriefing questionnaire

In the first part of the debriefing questionnaire, most participants (50.41%) indicated that the most likely research aim of the study was to investigate the benefits of reading. On average, learners in the form-focused group further indicated that they rarely ($M = 2.22$, $SD = 2.50$) noticed the enhancement via multiple-choice options. They also considered that the enhancement helped them little to pay attention to the use of English phrasal verbs while they were reading ($M = 1.90$, $SD = 2.15$), and that they felt little improvement in using English phrasal verbs after reading the forty texts ($M = 1.87$, $SD = 2.04$). Regarding how the enhancement impacted their reading, learners mostly reported that it had no impact at all (40%), while others indicated that it made their reading easier (20.89%).

In the second part of the debriefing questionnaire containing questions related to participants' technology background, all participants indicated that they owned a computer, and that, on average, they had been using a computer for more than ten years ($M = 12.10$, $SD = 4.05$).

Participants also reported spending an average of 2.29 hours per day (*SD* = 1.62) reading news websites, e-books, research articles, and social media. When asked where they used a computer, most indicated that did it in various locations (57.89%), including at home and at the university. When queried whether they had participated in non-traditional learning formats, the majority of participants (72.36%) revealed that they had not done so; those who did had an average of 3.31 (*SD* = 4.05) months of experience. Most learners (69.92%) also reported not having taken any online courses before. Finally, participants particularly liked the broad range of topics in the forty texts and greatly disliked that there was no report of reading progress during the experiment.

6.5 Summary

This chapter reported and summarized the results of the analyses on the study data. Overall, the results suggest that learning was similar in both treatment conditions. Results also reveal that individual differences in working memory interacted with the form-focused condition to predict learning gains. With respect to declarative memory, only nonverbal declarative memory was related to learning. Further, findings indicate that the development of accuracy in the multiple-choice activities during over-two-week reading period was nonlinear, and that this accuracy was not significantly associated with any of the cognitive abilities under investigation. Moreover, the effect of input/output enhancement on global reading comprehension was correlated with learners' reading proficiency level: intermediate learners performed lower than advanced learners. The discussion of these results will be presented in the next chapter.

Chapter 7

Discussion

This chapter discusses the outcomes of the present dissertation in relation to the three research questions investigated. To that end, it revisits each research question and discusses whether the predictions were supported or not. It concludes with a discussion of the results from the analysis of the learner log data collected during treatment, which were possible to gather thanks to the unique methodological affordances of using an ICALL system for SLA research.

The first research question examined whether working memory and declarative memory had an impact on the acquisition of English phrasal verbs. In the case of working memory, it was predicted that this individual difference would have an influence on learning outcomes. The results showed that working memory impacted L2 lexical development, indicating that the prediction made was thus confirmed. These findings are consistent with previous research which has documented working memory as an important predictor of L2 learning (e.g., Jeon & Yamashita, 2014; Kempe et al., 2009; Linck et al., 2014; Linck & Weiss, 2015; Malone, 2018; Martin & N. C. Ellis, 2012; J. N. Williams, 2012; Yang et al., 2017). It is worth pointing out that the positive effect of working memory was statistically the same for both receptive and productive lexical abilities, which is in line with meta-analytic research which has likewise reported that working memory is positively associated with the development of both receptive and productive L2 skills (e.g., Jeon & Yamashita, 2014; Linck et al., 2014). However, the positive impact of working memory on L2 outcomes was dependent on the instructional context. More specifically, working memory was positively related to vocabulary development only in the form-focused condition, suggesting the presence of an aptitude-treatment interaction (Cronbach & Snow, 1977; Snow, 1992; see below).

In regard to declarative memory, it was hypothesized that it was likely to have an effect on L2 lexical learning. The results showed that declarative memory was predictive of vocabulary acquisition, indicating that the prediction was therefore supported. These results follow the theoretical predictions made by the declarative/procedural model (Ullman, 2001, 2004, 2015, 2016), which posits a role for declarative memory in the acquisition of L2 word knowledge, including idiosyncratic knowledge at the multi-word level, as it is the case of phrasal verbs (Buffington & Morgan-Short, in press; Hamrick et al., 2018; Morgan-Short & Ullman, 2012; Ullman, 2015, 2016; Ullman & Lovelett, 2018). However, the relationship between declarative memory and learning was contingent on the type of declarative memory being indexed, i.e., not significant for verbal declarative memory, as assessed by the MLAT5, but significant for nonverbal declarative memory, as measured by the CVMT. And, just as for working memory, this relationship was different depending on the instructional condition, with a nonsignificant negative effect, or no effect at all, on the meaning-focused group, and a positive effect that approached significance in the form-focused group. This finding could be taken as a suggestion that there was a beneficial effect of nonverbal declarative memory in the form-focused group, but not in the meaning-focused group. Nonetheless, this finding needs to be replicated in future research.

On the whole, the findings suggest that the relationship between declarative memory and L2 word learning is complex. As noted in Chapter 2, the theoretical predictions concerning the role of declarative memory on L2 lexical acquisition are based on theories on declarative memory as a general-purpose learning system and research on child language acquisition (Buffington & Morgan-Short, in press; Hamrick et al., 2018). Therefore, strict interpretations of these predictions may not be as straightforward to be applied. Different factors, internal (e.g., proficiency, gender) and external (e.g., learning context), are likely to strengthen or weaken the potential effect on declarative memory on lexical development (Carpenter, 2008; Hamrick, 2015; Morgan-Short et al., 2014; Poldrack & Packard, 2003; Ullman, 2005, 2016). Hence, further research is warranted to investigate, for instance, whether learners' proficiency level could affect the predictive role of declarative memory in L2 vocabulary attainment. This echoes the call for more research made by Buffington and Morgan-Short (in press), who argue for more work to fully understand the role of declarative memory in L2 learning. On the other hand, Buffington and Morgan-Short also suggest that the type of task used to assess declarative memory might potentially affect the findings, as the effect of declarative memory might be somewhat dependent on the cognitive task employed. As indicated above, in this dissertation, only the scores of the nonverbal declarative task (CVMT) and not those of the verbal learning task (MLAT5) were related to learning gains. Therefore, as Buffington and Morgan-Short recommend, more extension and replication work is clearly warranted.

The second research question explored whether the potential impact of working memory and declarative memory was modulated by instruction type (meaning-focused versus form-focused). Regarding working memory, based on previous research, it was expected that the effect of working memory was to be modulated by the instructional condition (Ando et al., 1992; Denhovska & Serratrice, 2017; Indrarathne & Kormos, 2018; Malone, 2018; Tagarelli et al., 2015; Yang et al., 2017), and that the impact would more likely occur in the form-focused group (Ando et al., 1992; Indrarathne & Kormos, 2018; Malone, 2018; Tagarelli et al., 2015; Yang et al., 2017). In particular, participants with higher working memory capacity were hypothesized to benefit more from form-focused instruction (Brooks et al., 2006; Combs, 2005; Doughty, 2001; Ellis, 2012; Linck & Weiss, 2015; Meguro, 2017; Robinson, 1995; Sagarra, 2017; Sanz et al., 2016; J. N. Williams, 2012). The results revealed a significant interaction between working memory and the instructional condition, indicating that the influence of working memory on phrasal verb learning depended on the learning context. Further analyses showed that working memory was related to learning only in the form-focused group, suggesting an aptitude-treatment effect (Cronbach & Snow, 1977; Ellis, 2012; Jonassen & Grabowski, 1993; Kormos, 2013; Larsen-Freeman, 2009; Mackey, 2017; Roehr, 2012; Snow, 1991, 1992; Vatz et al., 2013).

The results matched the predictions of the study regarding the interaction between individual cognitive abilities and different types of instruction (Jonassen & Grabowski, 1993; Vatz et al., 2013), thereby supporting previous literature on aptitude-treatment interaction which has shown that the role of working memory in L2 learning varies depending on instructional conditions (Ando et al., 1992; Brooks et al., 2006; Denhovska & Serratrice, 2017; Erlam, 2005; Goo, 2012; Indrarathne & Kormos, 2018; Lado, 2017; Malone, 2018; Robinson, 2002; Sanz et al., 2016; Suzuki & DeKeyser, 2017; Tagarelli et al., 2015; Yang et al., 2017; Yilmaz, 2013).

These results may be explained by the following two reasons. First, the form-focused condition was more explicit in nature than the meaning-focused condition, in the sense that it drew learners' attention to linguistic form (de Graaff & Housen, 2009; DeKeyser, 1995; Ellis, 1994, 2012; Goo et al., 2015; Hulstijn, 2005; Indrarathne, Ratajczak, & Kormos, 2018; Norris & Ortega, 2000; Robinson et al., 2012; Schmidt, 1990, 2001, 2010; Sharwood Smith, 1991, 1993, 2013). Working memory has been shown to be more strongly associated to explicit processes (Baars, 1993; Erçetin & Alptekin, 2013; Li, 2017a; Santamaria & Sunderman, 2015). And second, as opposed to the meaning-focused condition, learners in the form-focused group were engaged in a higher attentional demanding task, as they had to attend to the meaning of the text and complete the multiple-choice activities simultaneously. According to Robinson (2001b, p. 29), the cognitive complexity of a given task is determined by "the attentional, memory, reasoning, and other information processing demands imposed" on language learners. Research has indicated that the more cognitively taxing the task (i.e., higher cognitive load), the more working memory resources are drawn upon to successfully complete it (Brooks et al., 2006; Combs, 2005; Daneman & Carpenter, 1980; Doughty, 2001; Ellis, 2012; Engle, Kane, & Tuholski, 1999; Fontanini & Tomitch, 2009; Li, 2017a; Linck & Weiss, 2015; Meguro, 2017; Robinson, 1995, 2011; Sagarra, 2017; Sanz et al., 2016; J. N. Williams, 2012).

It is noteworthy to mention that 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., Kempe, Brooks, & Kharkhurin, 2010; Tagarelli et al., 2015), but in a more realistic setting. Therefore, this study provides further evidence of the predictive validity of working memory for L2 learning in more ecologically relevant contexts (Linck & Weiss, 2015).

Given that declarative memory has been linked to more explicit learning processes (Poldrack & Packard, 2003; Ullman, 2016; Ullman & Lovelett, 2018), it was expected that the effect of declarative memory on lexical learning would be affected by the instructional context. More specifically, the effect was predicted to be present in the form-focused group, the more explicit instructional condition in the study. Recall that the form-focused group was asked to pay attention to linguistic form, as they were instructed to notice the blanks next to verbs and to click on the dropdown menus to complete the blanks in the texts. Although there was nonsignificant results for any of the two instructional conditions individually, there was a positive trend in the form-focused condition, suggesting a beneficial effect of declarative memory in more explicit instructional contexts supported by ICALL, and more generally, in L2 instruction where learners' attention is drawn to linguistic form (Robinson et al., 2012).

The third research question investigated whether an ICALL environment could provide the affordances necessary to conduct large scale, experimental-style research addressing the research questions of the present dissertation. The discussion of the research questions presented above clearly indicate that it was possible to collect experimental L2 data by means of a web-based ICALL system, as it was expected. The results suggest that 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 & Dickinson, 2017).

One of the methodological advantages offered by the use of ICALL systems for SLA research is the possibility of collecting data in order to explore the incremental process of learning during treatment (Meurers & Dickinson, 2017; Ziegler et al., 2017). To that end, as in Ziegler et al.'s (2017) pilot study, learner log data were also collected and analyzed in the current investigation. Particularly, these data afforded the opportunity to i) analyze the development of accuracy in the form-focused condition, and ii) assess the potential three-way interaction between cognitive individual differences, reading proficiency in English, and instructional conditions in predicting global reading comprehension of the experimental texts. In the case of the latter, this second-order interaction (i.e., the interaction of three variables; Elliott, 1998) was included because there has been little research on these types of interactions in SLA (DeKeyser, 2012b; but see e.g., Benson & DeKeyser, 2018; Russell, 2014; Tagarelli et al., 2016). It will be recalled that learners in the form-focused condition read and completed multiple-choice activities as a form of automatic input/output enhancement (Izumi, 2002; Meurers et al., 2010; Sharwood Smith, 1991, 1993; Swain, 1985, 1995), and that learners in both instructional conditions answered a main idea question after reading each text as a measure to test global reading comprehension (see Chapter 5). The following is a discussion of the analysis of performance during treatment.

In regard to the development of accuracy during treatment in the form-focused condition, the results indicated that this development was not linear, but parabolic, or U-shaped, 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 (DeKeyser, 1997; Huebner, 1983; Lightbown, 1983; Murakami, 2016; N. C. Ellis & Schmidt, 1998), thereby coinciding with previous research which has found evidence for nonlinearity in L2 development (Bulté & Housen, 2018; Murakami, 2016; Verspoor, Lowie, & Van Dijk, 2008). Another possible 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 & Housen, 2018). It is important to note, however, that the results of the feedback questionnaire (see Chapter 6) showed that, on average, learners mostly enjoyed reading the experimental texts (3/5, in the scale). Finally, it could be the case that learners were presented with less complex texts in the first half of the experiment (20th-21st texts) than in the second half, which could explain why accuracy first increased and then decreased. Nonetheless, again, learners reported that the texts were generally not very difficult to read (3/5, in the scale). As indicated in Chapter 5, the texts were shown in fixed order, which could have introduced a difficulty bias. Future research should consider randomizing the order of the experimental texts so as to reduce the potential influence of differences in text difficulty.

Additionally, and considering that most learners were at ceiling in accuracy throughout the experiment when doing the multiple-choice activities, a further analysis with only those who steadily scored under 95% (n = 20) was conducted. The results revealed that accuracy in this subgroup of learners significantly decreased with time. Similar findings were reported by Ziegler et al. (2017), who found that accuracy was negatively associated with time in the multiple-choice group. Just as in Ziegler et al.'s case, the present study did not aim to isolate the effect of selecting a particular option from different alternatives. In future studies, it would be advisable to investigate the possible effects of isolating the response options in the multiple-choice format. However, given that the overall aim of the present dissertation is to explore the role of individual differences in L2 vocabulary acquisition in a web-based learning environment, this is not further discussed herein.

In regard to the three-way relationship between individual differences, reading proficiency, and instructional conditions, the results indicated that these factors did not interact to predict general understanding of the experimental texts. However, the results showed that reading proficiency and instructional conditions interacted to account for learners' comprehension, meaning that the effect of the pedagogical context was a function of learners' reading proficiency level. Further analyses indicated that reading comprehension was negatively impacted by the instructional condition but only for learners at the intermediate level. Intermediate learners in the form-focused group, in particular, scored lower, on average, than those of the same proficiency level in the meaning-focused condition. In other words, for intermediate 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. These results are in line with previous research which has found a trade-off for simultaneous processing of form and meaning (Boers, Warren, et al., 2017; S. Choi, 2017; Greenslade, Bouden, & Sanz, 1999; S.-K. Lee, 2007; Overstreet, 1998; E. S. Park & Nassif, 2014; H. Park et al., 2012; VanPatten, 1990). Moreover, these findings also align with the idea that the (unfavorable) effects of paying attention to form and meaning simultaneously is related to language experience (Greenslade et al., 1999; Leow et al., 2008; Morgan-Short, Heil, Botero-Moriarty, & Ebert, 2012; VanPatten, 1990). Particularly, learners at lower levels of proficiency are more likely to have difficulty simultaneously attending to form and meaning because they need to devote more attentional resources to making form-meanings connections than more advanced learners (Leow et al., 2008; VanPatten, 1990).

In regard to the overall effect of individual differences, the results revealed that neither working memory nor declarative memory statistically affected reading comprehension during treatment. In the case of the former, these findings parallel those of previous studies that have reported that working memory failed to account for learners' comprehension (e.g., S. Choi, 2013; D. M. Chun & Payne, 2004). A possible reason for this might be that only inferential reading comprehension, that is, comprehension skills at the macro-level (i.e., main idea; Grabe, 2009; Kintsch & Van Dijk, 1978) was measured. For instance, S. Choi (2013) investigated the role of working memory in predicting both inferential and local comprehension, with the latter being understood as comprehension skills at the micro-level (e.g., details). S. Choi's (2013) study found that working memory was only predictive of local, not inferential comprehension. S. Choi explains that learners "may not have been [cognitively] overloaded during inference generation" (p. 37). This might have also been the case for learners in the present study. Another reason could be that assessing reading comprehension using almost exclusively one multiple-choice question (i.e., "What is the main idea of the article?") was not taxing enough for working memory. To investigate the potential contribution of working memory in the context of web-based reading, further research should include different techniques (e.g., cloze tasks) to gauge both macro-level and micro-level reading skills (see Koda, 2012). Overall, the findings suggest that working memory might not be a significant predictor of reading comprehension when learners read news texts on the Internet (D. M. Chun & Payne, 2004; Joh & Plakans, 2017).

Regarding declarative memory, the declarative/procedural model (Ullman, 2001, 2004, 2015, 2016) does not postulate a role for declarative memory in L2 reading comprehension, and therefore its inclusion in the interaction was exploratory. This means that the results should primarily be used to inform future investigations. However, it can be speculated that declarative memory might be related to reading skills in the L2 because of its assumed underlying role in lexical knowledge (Ullman, 2001, 2004, 2015, 2016). L2 vocabulary knowledge is one of the most important predictors of L2 reading comprehension (Jeon & Yamashita, 2014). On the other hand, it has been suggested that, albeit different, declarative memory and working memory are closely associated with one another (Lum et al., 2012). For instance, working memory has been indicated to contribute to organizing information before it becomes stored in the declarative memory (Fletcher, Shallice, Frith, Frackowiak, & Dolan, 1998; Lum et al., 2012). It follows that, since working memory is also an important correlate of L2 reading comprehension (Jeon & Yamashita, 2014), it is plausible that declarative memory might as well emerge as a predictor of learners' comprehension.

In sum, the present chapter discussed the results in light of the three research questions that guided the present study. The results reveal that both working memory and declarative memory were predictive of L2 lexical learning. However, only the working memory effect was modulated by the instructional context, with the effect being found exclusively in the form-focused condition. The chapter also presented a discussion of the learners' performance during treatment. This was possible due to the methodological affordances available within an ICALL system. The next chapter discusses the general aspects, limitations, and implications of the study.

Part III

Chapter 8

General discussion

Considering the present study's findings, this chapter discusses first i) the role of cognitive individual differences in ICALL-supported learning environments, and then ii) the relevance of conducting research at the intersection of ICALL and SLA. This is followed by a discussion of the limitations of the study, directions for future research, and pedagogical implications of the findings.

8.1 Cognitive individual differences and ICALL

The current study examined the relationship between cognitive individual differences, namely working memory and declarative memory, and L2 vocabulary acquisition in a web-based ICALL environment. Regarding working memory, research has found that it plays a role in L2 development (e.g., Doughty, 2001; Li, 2017a; Linck et al., 2014; J. N. Williams, 2012). Results from the current study suggest a differential role of working memory in the development of phrasal verb knowledge under different instructional conditions. In particular, the effect of working memory was only evident in the form-focused group. These findings lend support to the argument that factors such as context can influence the effect of working memory on L2 acquisition (e.g., Faretta-Stutenberg & Morgan-Short, 2018; Indrarathne & Kormos, 2018; Lado, 2017; Linck et al., 2014; Malone, 2018; Sanz et al., 2016; Tagarelli et al., 2015; 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 (Brooks et al., 2006; Combs, 2005; Daneman &

Carpenter, 1980; Doughty, 2001; Ellis, 2012; Engle et al., 1999; Fontanini & Tomitch, 2009; Li, 2017a; Linck & Weiss, 2015; Meguro, 2017; Robinson, 1995, 2011; Sagarra, 2017; Sanz et al., 2016; J. N. Williams, 2012), as well as ii) previous research that has found that working memory interacts with instructional treatments (Ando et al., 1992; Brooks et al., 2006; Denhovska & Serratrice, 2017; Erlam, 2005; Goo, 2012; Indrarathne & Kormos, 2018; Lado, 2017; Malone, 2018; Robinson, 2002; Sanz et al., 2016; Suzuki & DeKeyser, 2017; Tagarelli et al., 2015; Yang et al., 2017; Yilmaz, 2013). It may be concluded that working memory seems to be particularly important in ICALL-mediated interventions whereby learners' attention is drawn to linguistic form. Regarding declarative memory, it has been posited to play a role in L2 development as well (Ullman, 2015, 2016). Of significance here, theoretical considerations suggest that declarative memory may underlie the acquisition of lexical knowledge, including the acquisition of multi-word lexical units (Ullman, 2016; Ullman & Lovelett, 2018). The results of the present study reveal a role for declarative memory in phrasal verb acquisition, as suggested by theoretical predictions (Ullman, 2015, 2016). However, as noted in the previous chapter, the impact of declarative memory on learning was dependent on the kind of declarative memory being measured, which in this case it was only significant for nonverbal declarative memory. More generally, it could be assumed that the association between declarative memory and L2 learning is complex, as many different factors are likely to be involved. In this regard, Ullman (2015) concludes:

> [a]ccording to the DP [declarative/procedural] model, both behavioral and neural correlates of L2 learning should vary on the basis of multiple factors, including biological variables (e.g., sex and genetic variability), input variables (e.g., amount and type of L2 exposure), and linguistic subsystems (e.g., lexicon vs. grammar). Moreover, a number of these variables likely *interact* [emphasis added] (p. 156).

In general, the results of this investigation are in line with the overall findings of individual difference research, in that learners' cognitive differences are crucial predictors of L2 acquisition (e.g., Buffington & Morgan-Short, in press; Hamrick et al., 2018; Indrarathne & Kormos, 2018; Malone, 2018; Robinson, 1995; Tagarelli et al., 2015, 2016). Thus, they should be taken into account when explaining differential success in learning another language. Moreover, the results also support the general findings of aptitude-treatment interaction research, in the sense that individual differences in cognitive abilities are of greater importance when the learning burden is increased (Benson & DeKeyser, 2018; Brooks et al., 2006; Sanz et al., 2016; Suzuki & DeKeyser, 2017).

8.2 Relevance of ICALL

Collaborative research at the intersection of ICALL and SLA is still quite scarce. Nonetheless, NLP-driven ICALL systems have been identified as very promising tools to support language learning and to investigate L2 development (Meurers & Dickinson, 2017; Ziegler et al., 2017). These systems have the potential to offer the necessary affordances to collect empirically controlled data on the processes and outcomes of L2 learning and the results of the current dissertation nicely illustrate these affordances. Building on the work by Ziegler et al. (2017), the capabilities of the WERTi system (Meurers et al., 2010) were expanded here. In comparison to Ziegler et al.'s study whereby the WERTi system was only used to administer the instructional treatments, an updated version of this web-based ICALL system was employed in the present investigation to systematically implement treatments and testing, including the testing of individual differences. In the following discussion, the results of the present study serve as a backdrop to highlight some of the several analytical and methodological advantages ICALL systems have to offer to language learning and SLA research (see Chapter 2).

In the current study, the integration of individual difference tests, language tests, and instructional treatments in a web-based ICALL system provided the opportunity to gather empirical evidence on how learner individual variables may moderate the effects of instruction—the products of learning. As explained before, the results indicate that individual differences in working memory were related to vocabulary gains, whereas individual differences in declarative memory were not. Importantly, the findings suggest that the benefits of enhancing L2 input as a pedagogical intervention may be influenced by individual differences in cognitive abilities, that is, an aptitude-treatment interaction (Cronbach & Snow, 1977; Jonassen & Grabowski, 1993; Larsen-Freeman, 2009; Snow, 1991, 1992; Vatz et al., 2013). Moreover, considering the inconsistent and contradictory findings on the efficacy of visual input enhancement (S.-K. Lee & Huang, 2008; Leow & Martin, 2018), the results reported here can contribute to elucidating the role of input/output enhancement in L2 development by looking at the interaction between individual differences and instruction, as suggested by several SLA researchers (e.g., DeKeyser, 2012b; Spada, 2011). With respect to the process of learning, the logging of learners' interaction while using the ICALL system allowed to determine that i) the learning *during* treatment in the form-focused group was nonlinear (i.e., path of acquisition; Murakami, 2013), and ii) reading comprehension in the intermediate, not advanced level of proficiency was hindered by the simultaneous processing of form and meaning. The data to explore these classical SLA topics (i.e., aptitude-treatment interactions, input/output enhancement, simultaneous attention to form and meaning) were obtained thanks to possibilities offered by an ICALL system that not only provided a platform for web-based language learning but also for Internet testing.

Using a web-based ICALL system in the current research also afforded some further methodological advantages. For instance, in contrast to lab-based research, it was possible to investigate instructed SLA in everyday contexts (e.g., home, university), as the ICALL system kept track of learners' interaction during real-life language learning (Ziegler et al., 2017). Likewise, learners were exposed to authentic materials in the form of news articles on the web, which were automatically enhanced on the fly (Meurers et al., 2010), meaning that the exposure to target linguistic form occurred in an ecologically valid environment. Moreover, the web-based ICALL system as well allowed the collection of data from a greater number of participants and in a faster fashion, which would not be otherwise feasible to accomplish in the lab or in the classroom. This can be attributable to the reduction of logistical issues related to lab- or classroom-based research, such as coming to the lab or attending a class, or returning for multiple testing sessions (Ziegler et al., 2017). To illustrate this point, the data in this study was collected from 127 participants during an approximate two-week treatment period, with the experiment time of each participant totaling around eight hours (see Chapter 5). Had this been done in the lab, for example, the data collection time would have amounted to around 4.5 months. With respect to the study's sample size, this was much more larger than the typical median group size in L2 research, which has been indicated to be 19 participants (Plonsky, 2013). A larger sample size is important because it increases statistical power, which, in turn, allows for more generalizability of findings (Krantz & Reips, 2017; Plonsky, 2017).

Overall, ICALL systems hold the promise of assisting researchers in understanding L2 development more deeply and in a scalable way. Although web-based experimentation is still an underdeveloped area in SLA research (MacWhinney, 2017), ICALL-mediated studies such as this can certainly contribute to investigating language learning beyond conventional learning and experimental environments.

8.3 Limitations of the study

There are some limitations to this dissertation. The first limitation is the advanced proficiency of the participants. Although phrasal verbs are notoriously difficult for English learners, even at advanced levels (Cornell, 1985; Garnier & Schmitt, 2015, 2016; Larsen-Freeman & Celce-Murcia, 2015; Yasuda, 2010; see Chapter 3), it is possible that the proficiency level was too high for significant differences between the two instructional treatments (meaning-focused vs. form-focused instruction) to emerge. In the case of form-focused instruction, learner proficiency is regarded as one of the most important factors to influence the impact of this type of instruction (Benati, 2016; S.-K. Lee & Huang, 2008; J. Williams, 1999). In visual input enhancement research, in particular, intermediate learners are typically the target learning group, since they are thought to have the necessary skills to process the visually enhanced reading materials; advanced learners, on the other hand, are believed to have much more exposure to targeted linguistic forms, which might be an important factor in establishing the effectiveness of input enhancement (S.-K. Lee & Huang, 2008). Given the web-based nature of the present investigation, however, participant recruitment was as inclusive as possible in order to account for the tendency for high attrition (dropout) rates associated with web-based research (Krantz & Reips, 2017; Reips, 2002), thereby reducing potential problems of missing data.

A second limitation is the lack of a test-only control group. This type of control group is used to account for possible testing effects, as learners might learn from taking the tests twice (Bardovi-Harlig, 2015). However, had this been the case in the current study, the testing effect would have been seen in both treatment groups, but this was not found in the statistical analyses. Nonetheless, the inclusion of test-only controls could have ruled out this possibility entirely.

Finally, a further limitation is related to the absence of a delayed posttest. This test would have allowed to determine whether the findings regarding the superior benefits of form-focused instruction for learners with higher working memory capacity would be enduring over time (Ando et al., 1992; Erlam, 2005; Mackey, Philp, Egi, Fujii, & Tatsumi, 2002). Unlike immediate posttests, delayed tests are thought to be capable of capturing the full extent of learning, as subsequent memory consolidation processes are likely to take place once instructional treatments have been completed (Sanz & Grey, 2015; see also Ellis, 2012).

8.4 Directions for future research

To the researcher's knowledge, this is the first study to use an ICALL system to conduct a web-based experiment in L2 research that integrates both testing and treatments into one platform. Consequently, replications of the current study are warranted.

Future research should also strive to investigate learners with lower proficiency than those who participated in this study. As explained in the limitation section, intermediate-level learners should be the target group, as this proficiency level could be more adequate to shed further light on the potential influence of individual differences in web-based ICALL-supported contexts.

Another suggestion for future research is to include other individual difference variables such as language aptitude and personality. This inclusion would allow for a greater characterization of those learners who are more likely to benefit from ICALL interventions. In a similar vein, it would be of interest to allow for different individual differences interact among themselves and other potential factors, such as learner proficiency and the nature of the target form (see DeKeyser, 2012b, for arguments supporting interactional research in SLA). To run these interaction analyses, it would be necessary to obtain sufficient amount of data to achieve statistical power (Cohen, 1988). Given the methodological affordances offered by web-based ICALL systems, collecting data from larger sample sizes in future studies should not be a problem.

Lastly, additional investigations should include other instructional conditions such as visual enhancement via fill-in-the-blanks activities. In this regard, it would be interesting to further explore which pedagogical activities are more conducive to learning in ICALL-mediated pedagogical treatments (see Ziegler et al., 2017).

8.5 Pedagogical implications

The present study has implications for L2 pedagogy using web-based ICALL systems as well as instructed SLA in general. The results have shown that individual differences in working memory mediate the effects of form-focused instruction in web-based ICALL interventions. These findings suggest that high working memory capacity contributes to the learning of English phrasal verbs, which, as noted earlier, are notoriously challenging forms for learners to master. More generally, the results suggest that individual differences in cognitive abilities are predictive of L2 attainment in a web-based learning setting and therefore should be taken into consideration when implementing instructional treatments in this context.

In the case of the visual input/output enhancement technique implemented in this ICALL study, namely multiple-choice options, the findings suggest that this type of form-focused instruction is more beneficial for L2 lexical development, insofar as learners' working memory capacity is large enough to attend to both meaning and form simultaneously. For learners with fewer working memory resources, however, more support would be needed. For instance, scaffolding feedback could be given to these learners when they encounter difficulty (see, e.g., Rudzewitz et al., 2017).

Finally, having information about learners' profiles could be used to match instruction to their strengths and weaknesses (Cronbach & Snow, 1977; Jonassen & Grabowski, 1993). While collecting this information via testing in the classroom may be impractical, the findings from the present study show that this is possible to accomplish in a web-based environment. Therefore, the findings suggest that future pedagogical interventions employing ICALL systems could further support L2 learning by tailoring instruction to learners' particular characteristics. For example, based on the results of individual difference measures administered at the beginning of the intervention, ICALL systems could, accordingly, provide learners with different types of form-focused instruction (Doughty & Williams, 1998; Ellis, 2016).

Chapter 9

Conclusion

The current study sought to investigate i) whether individual differences in working memory and declarative memory would affect the learning of L2 vocabulary in the context of a large-scale, web-based ICALL intervention, ii) whether the potential effect of working memory and declarative memory would be modulated by instructional conditions, and iii) whether an ICALL system would support the affordances required for answering the prior research questions.

Regarding the products of learning, this study provided evidence to suggest that both working memory and declarative memory influenced the acquisition of lexical knowledge. However, only the influence of working memory was modulated by instructional condition, with evidence for an aptitude-treatment interaction in the form-focused instructional condition. As for the process of learning, some evidence was observed that development during form-focused instruction was not linear, and that learner proficiency played a role on reading comprehension, with intermediate learners experiencing a trade-off between meaning and form.

In sum, the present study clearly illustrates the advantages of conducting interdisciplinary work between ICALL and SLA, as empirical evidence of the role of individual differences as well as the process and outcomes of L2 acquisition was collected through an all-encompassing ICALL system. As such, it also constitutes the first attempt to characterize who is likely to benefit from ICALL-mediated interventions.

References

Ahmadian, M. J. (2012). The relationship between working memory capacity and L2 oral performance under task-based careful online planning condition. *TESOL Quarterly*, *46*(1), 165–175.

Ai, H. (2017). Providing graduated corrective feedback in an intelligent computer-assisted language learning environment. *ReCALL*, *29*(3), 313–334. https://doi.org/10.1017/S095834401700012X

Aiken, L. S., & West, S. G. (1991). *Multiple regression: Testing and interpreting interactions*. Newbury Park, CA: SAGE.

Alanen, R. (1995). Input enhancement and rule presentation in second language acquisition. In R. W. Schmidt (Ed.), *Attention and awareness in foreign language learning* (pp. 259–302). Honolulu: University of Hawai'i, Second Language Teaching & Curriculum Center.

Alcón Soler, E. (2005). Does instruction work for learning pragmatics in the EFL context? *System*, *33*(3), 417–435. https://doi.org/10.1016/j.system.2005.06.005

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

Alderson, J. C., & Huhta, A. (2005). The development of a suite of computer-based diagnostic tests based on the Common European Framework. *Language Testing*, *22*(3), 301–320. https://doi.org/10.1191/0265532205lt310oa

Alderson, J. C., Nieminen, L., & Huhta, A. (2016). Characteristics of weak and strong readers in a foreign language. *The Modern Language Journal*, *100*(4), 853–879.

https://doi.org/10.1111/modl.12367

Alejo-González, R. (2010). Making sense of phrasal verbs: A cognitive linguistic account of L2 learning. *AILA Review*, *23*(1), 50–71. https://doi.org/10.1075/aila.23.04ale

Alejo-González, R. (2012). Variability in L2 acquisition of formulaic sequences: A study of linguistic factors in the use and learning of phrasal verbs by non-native speakers. *Language Value*, *4*(1), 33–62. https://doi.org/10.6035/LanguageV.2012.4.3

Allen, R. J., Hitch, G. J., & Baddeley, A. D. (2018). Exploring the sentence advantage in working memory: Insights from serial recall and recognition. *Quarterly Journal of Experimental Psychology*. https://doi.org/10.1177/1747021817746929

Alptekin, C., Erçetin, G., & Özemir, O. (2014). Effects of variations in reading span task design on the relationship between working memory capacity and second language reading. *The Modern Language Journal*, *98*(2), 536–552. https://doi.org/10.1111/modl.12089

Alsadoon, R., & Heift, T. (2015). Textual input enhancement for vowel blindness: A study with Arabic ESL learners. *The Modern Language Journal*, *99*(1), 57–79. https://doi.org/10.1111/modl.12188

Amaral, L. A., & Meurers, D. (2011). On using intelligent computer-assisted language learning in real-life foreign language teaching and learning. *ReCALL*, *23*(1), 4–24.

American Psychological Association. (1986). *Guidelines for computer-based tests and interpretations*. Washington, DC: American Psychological Association.

Ando, J., Fukunaga, N., Kurahachi, J., Suto, T., Nakano, T., & 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.

Antoniadis, G., & Desmet, P. (2016). NLP for learning and teaching: Challenges and opportunities. *Traitement Automatique des Langues*, *57*(3), 7–13.

Antoniou, M., Ettlinger, M., & Wong, P. C. (2016). Complexity, training paradigm design, and the contribution of memory subsystems to grammar learning. *PLoS ONE*, *11*(7), e0158812. https://doi.org/10.1371/journal.pone.0158812

Arispe, K. (2012). *Why vocabulary still matters: L2 lexical development and learner autonomy as mediated through an ICALL tool, Langbot* (Doctoral dissertation). University of California, Davis, Davis, CA.

Armstrong, K. (2004). Sexing up the dossier: A semantic analysis of phrasal verbs for language teachers. *Language Awareness*, *13*(4), 213–224. https://doi.org/10.1080/09658410408668808

Azab, M., Salama, A., Oflazer, K., Shima, H., Araki, J., & Mitamura, T. (2013). An NLP-based reading tool for aiding non-native English readers. In *Proceedings of the International Conference Recent Advances in Natural Language Processing* (pp. 41–48). Hissar, Bulgaria.

Azzaro, G. (2012). Phrasal verbs through DDL. *Ricerche di Pedagogia e Didattica. Journal of Theories and Research in Education*, *7*(2), 1–24. Retrieved from https://rpd.unibo.it/article/view/3217/2603

Baars, B. J. (1993). *A cognitive theory of consciousness*. New York: Cambridge University Press.

Baayen, R. H. (2008). *Analyzing linguistic data: A practical introduction to statistics using R*. Cambridge: Cambridge University Press.

Baayen, R. H., Davidson, D., & Bates, D. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, *59*(4), 390–412. https://doi.org/10.1016/j.jml.2007.12.005

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

Baddeley, A. D. (2003). Working memory and language: An overview. *Journal of Communication Disorders*, *36*(3), 189–208. https://doi.org/10.1016/

S0021-9924(03)00019-4

Baddeley, A. D. (2007). *Working memory, thought, and action*. New York: Oxford University Press.

Baddeley, A. D. (2017). Modularity, working memory and language acquisition. *Second Language Research*, *33*(3), 299–311. https://doi.org/10.1177/0267658317709852

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

Bailey, H., Dunlosky, J., & Kane, M. J. (2011). Contribution of strategy use to performance on complex and simple span tasks. *Memory & Cognition*, *39*(3), 447–461. https://doi.org/10.3758/s13421-010-0034-3

Bailey, S. K. T., Neigel, A. R., Dhanani, L. Y., & Sims, V. K. (2017). Establishing measurement equivalence across computer- and paper-based tests of spatial cognition. *Human Factors*. https://doi.org/10.1177/0018720817747731

Bamford, J., & Day, R. R. (2004). *Extensive reading activities for teaching language*. Cambridge: Cambridge University Press.

Baranski, M. F. S., & Was, C. A. (2018). A more rigorous examination of the effects of mindfulness meditation on working memory capacity. *Journal of Cognitive Enhancement*. https://doi.org/10.1007/s41465-018-0064-5

Bardovi-Harlig, K. (2015). Designing instructional effect studies for L2 pragmatics: A guide for teachers and researchers. In S. Gesuato, F. Bianchi, & W. Cheng (Eds.), *Teaching, learning and investigating pragmatics: Principles, methods and practices* (pp. 135–164). Newcastle upon Tyne: Cambridge Scholars Publishing.

Bates, D., Kliegl, R., Vasishth, S., & Baayen, R. H. (2015). *Parsimonious mixed models.* Retrieved from http://arxiv.org/abs/1506.04967

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

Bates, E., Dale, P. S., & Thal, D. (1995). Individual differences and their implications for theories of language development. *The handbook of child language*, 96–151.

Behzadian, K. (2016). The effect of input enhancement and concordance-based activities on Iranian EFL learners' acquisition and retention of phrasal verbs. *Michigan Academician*, *44*(1), 29–50. https://doi.org/10.7245/0026-2005-44.1.29

Benati, A. G. (2016). Input manipulation, enhancement and processing: Theoretical views and empirical research. *Studies in Second Language Learning and Teaching*, *1*(6), 65–88. https://doi.org/10.14746/ssllt.2016.6.1.4

Benati, A. G. (2017). The role of input and output tasks in grammar instruction: Theoretical, empirical and pedagogical considerations. *Studies in Second Language Learning and Teaching*, *3*(7), 377–396. https://doi.org/10.14746/ssllt.2017.7.3.2

Benson, S., & DeKeyser, R. M. (2018). Effects of written corrective feedback and language aptitude on verb tense accuracy. *Language Teaching Research*. https://doi.org/10.1177/1362168818770921

Biber, D., Conrad, S., & Reppen, R. (1998). *Corpus linguistics: Investigating language structure and use*. Cambridge: Cambridge University Press.

Biber, D., Johansson, S., Leech, G., Conrad, S., & Finegan, E. (1999). *The Longman grammar of spoken and written English*. London: Longman.

Biedroń, A., & 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). https://doi.org/10.14746/ssllt.2016.6.3.3

Bird, S., Klein, E., & Loper, E. (2009). *Natural language processing with Python*. Sebastopol, California: O'Reilly.

Birjandi, P., Alavi, S. M., & Karimi, S. N. (2015). Effects of unenhanced, enhanced, and elaborated input on learning English phrasal verbs. *International Journal of Research Studies in Language Learning*, *4*(1), 43–59. https://doi.org/10.5861/ijrsll.2014.772

Bock, J. K. (1986). Syntactic persistence in language production. *Cognitive Psychology*, *18*(3), 355–387. https://doi.org/10.1016/0010-0285(86)90004-6

Boers, F. (2000). Metaphor awareness and vocabulary retention. *Applied Linguistics*, *21*(4), 553–571. https://doi.org/10.1093/applin/21.4.553

Boers, F. (2017). Intentional versus incidental learning. In J. I. Liontas (Ed.), *The TESOL encyclopedia of English language teaching.* John Wiley & Sons, Inc. https://doi.org/10.1002/9781118784235.eelt0074

Boers, F., Demecheleer, M., He, L., Deconinck, J., Stengers, H., & Eyckmans, J. (2017). Typographic enhancement of multiword units in second language text. *International Journal of Applied Linguistics*, *27*(2), 448–469. https://doi.org/10.1111/ijal.12141

Boers, F., Warren, P., Grimshaw, G., & Siyanova-Chanturia, A. (2017). On the benefits of multimodal annotations for vocabulary uptake from reading. *Computer Assisted Language Learning*, *30*(7), 709–725. https://doi.org/10.1080/09588221.2017.1356335

Boiteau, T. W., Malone, P. S., Peters, S. A., & Almor, A. (2014). Interference between conversation and a concurrent visuomotor task. *Journal of Experimental Psychology: General*, *143*(1), 295–311. https://doi.org/10.1037/a0031858

Bolinger, D. (1971). *The phrasal verb in English.* Cambridge, MA: Harvard University Press.

Bornstein, M. H., & Putnick, D. L. (2012). Stability of language in childhood: A multiage, multidomain, multimeasure, and multisource study. *Developmental psychology*, *48*(2), 477. https://doi.org/10.1037/a0025889

Boulton, A., & Cobb, T. (2017). Corpus use in language learning: A meta-analysis. *Language Learning*, *67*(2), 348–393. https://doi.org/10.1111/lang.12224

Brooks, P. J., Kempe, V., & Sionov, A. (2006). The role of learner and input variables in learning inflectional morphology. *Applied Psycholinguistics*, *27*(2), 185–209.

https://doi.org/10.1017/S0142716406060243

Buffington, J., & Morgan-Short, K. (2018). Construct validity of procedural memory tasks used in adult-learned language. In *The 40th Annual Meeting of the Cognitive Science Society.* Retrieved from http://mindmodeling.org/cogsci2018/papers/0276/0276.pdf

Buffington, J., & Morgan-Short, K. (in press). Declarative and procedural memory as individual differences in second language aptitude. In Z. Wen, P. Skehan, A. Biedroń, S. Li, & R. Sparks (Eds.), *Language aptitude: Multiple perspectives and emerging trends.* Routledge.

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

Bunting, M., & Engle, R. W. (2015). Foreword. In Z. Wen, M. Borges Mota, & A. McNeill (Eds.), *Working memory in second language acquisition and processing* (pp. xvii–xxiv). Bristol: Multilingual Matters.

Burns, M. K., Davidson, K., Zaslofsky, A. F., Parker, D. C., & Maki, K. E. (2017). The relationship between acquisition rate for words and working memory, short-term memory, and reading skills: Aptitude-by-treatment or skill-by-treatment interaction? *Assessment for Effective Intervention*. https://doi.org/10.1177/1534508417730822

Bywater, F. V. (1969). *A proficiency course in English*. London: University of London Press.

Capielo, C., Mann, L., Nevels, B., & Delgado-Romero, E. (2014). Multicultural considerations in measurement and classification of positive psychology. In T. Pedrotti, Jennifer, & L. M. Edwards (Eds.), *Perspectives on the intersection of multiculturalism and positive psychology* (pp. 31–44). Dordrecht: Springer Netherlands.

Cappelle, B., Shtyrov, Y., & 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. https://doi.org/10.1016/j.bandl.2010.09.004

Carpenter, H. (2008). *A behavioral and electrophysiological investigation of different aptitudes for L2 grammar in learners equated for proficiency level* (Doctoral dissertation). Georgetown University, Washington, DC..

Carrell, P. L. (1991). Second language reading: Reading ability or language proficiency? *Applied linguistics*, *12*(2), 159–179. https://doi.org/10.1093/applin/12.2.159

Carrell, P. L. (1992). Awareness of text structure: Effects on recall. *Language Learning*, *42*(1), 1–18. https://doi.org/10.1111/j.1467-1770.1992.tb00698.x

Carroll, J. B., & Sapon, S. M. (1959). *Modern Language Aptitude Test*. New York: The Psychological Corporation/Harcourt Brace Jovanovich.

Carroll, S. (2001). *Input and evidence: The raw material of second language acquisition*. Amsterdam: John Benjamins Publishing.

Carter, R., & McCarthy, M. (2006). *Cambridge Grammar of English: A comprehensive guide*. Cambridge: Cambridge University Press.

Chang, J. (2001). Chinese speakers. In M. Swan & B. Smith (Eds.), *Learner English: A teacher's guide to interference and other problems* (pp. 310–324). Cambrigde: Cambridge University Press.

Chen, M. (2013a). Overuse or underuse: A corpus study of English phrasal verb use by Chinese, British and American university students. *International Journal of Corpus Linguistics*, *18*(3), 418–442. https://doi.org/10.1075/ijcl.18.3.07che

Chen, M. (2013b). Phrasal verbs in a longitudinal learner corpus: Quantitative findings. In S. Granger, G. Gilquin, & F. Meunier (Eds.), *Twenty Years of Learner Corpus Research. Looking Back, Moving Ahead: Proceedings of the First Learner Corpus Research Conference (LCR 2011)* (pp. 89–101). Louvain-la-Neuve: Presses universitaires de Louvain.

Chen, X., & Smakman, D. (2016). Avoidance of phrasal verbs by learners of English: Definitional and methodological issues. In M. van Arnhem & D. Smakman (Eds.), *Proceedings Van Schools tot Scriptie II : een colloquium over universitair taalvaardigheidsonderwijs Universiteit Leiden, 4 juli 2014* (pp. 40–48). Retrieved from http://hdl.handle.net/1887/37680

Chévez-Herra, R. (2013). Phrasal verbs: Their teaching and acquisition. *Revista de Lenguas Modernas*, *19*, 487–509. Retrieved from http://hdl.handle.net/10669/23790

Chiang, Y.-T. (2012). Integrating input enhancement and processing instruction in promoting acquisition of English phrasal verbs. *The Asian EFL Journal Quarterly*, *14*(1), 10–46.

Chiang, Y.-T. (2013). *Effects of textual modifications on incidental phrasal verb acquisition through second language reading by Taiwanese learners of English* (Doctoral dissertation). Tamkang University, New Taipei City, Taiwan.

Chinkina, M., & Meurers, D. (2016). Linguistically aware information retrieval: Providing input enrichment for second language learners. In *Proceedings of the 11th Workshop on Innovative Use of NLP for Building Educational Applications BEA@NAACL-HLT 2016* (pp. 188–198). San Diego, California.

Cho, M. Y. (2010). The effects of input enhancement and written recall on noticing and acquisition. *Innovation in Language Learning and Teaching*, *4*(1), 71–87. https://doi.org/10.1080/17501220903388900

Choi, I.-C. (2016). Efficacy of an ICALL tutoring system and process-oriented corrective feedback. *Computer Assisted Language Learning*, *29*(2), 334–364. https://doi.org/10.1080/09588221.2014.960941

Choi, S. (2013). Working memory capacity, vocabulary knowledge, and reading comprehension of EFL learners. *English Language Teaching*, *25*(1), 25–42.

Choi, S. (2017). Processing and learning of enhanced English collocations: An eye movement study. *Language Teaching Research*, *21*(3), 403–426. https://doi.org/10.1177/1362168816653271

Chun, D. M., & Payne, J. S. (2004). What makes students click: Working memory and look-up behavior. *System*, *32*(4), 481–503. https://doi.org/10.1016/

j.system.2004.09.008

Chun, M. M. (2000). Contextual cueing of visual attention. *Trends in Cognitive Sciences*, *4*(5), 170–178. https://doi.org/10.1016/S1364-6613(00)01476-5

Cintrón-Valentín, M., & N. C. Ellis. (2015). Exploring the interface: Explicit focus-on-form instruction and learned attentional biases in L2 Latin. *Studies in Second Language Acquisition*, *37*(2), 197–235. https://doi.org/10.1017/S0272263115000029

Coe, N. (2001). Speakers of Spanish and Catalan. In M. Swan & B. Smith (Eds.), *Learner English: A teacher's guide to interference and other problems* (pp. 90–112). Cambrigde: Cambridge University Press.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Hillsdale, NJ: Erlbaum.

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

Combs, C. (2005). What cognitive processes are triggered by input enhancement. *Teachers College, Columbia University Working Papers in TESOL & Applied Linguistics*, *4*(3).

Comeaux, I., & McDonald, J. L. (2018). Determining the effectiveness of visual input enhancement across multiple linguistic cues. *Language Learning*, *68*(1), 5–45. https://doi.org/10.1111/lang.12255

Condon, N. (2008). How cognitive linguistic motivations influence the learning of phrasal verbs. In F. Boers & S. Lindstromberg (Eds.), *Cognitive linguistic approaches to teaching vocabulary and phraseology* (pp. 33–158). Berlin: Mouton de Gruyte.

Consigny, A. (2006). The polysemy (?) of phrasal verbs in English. *Word*, *57*(1), 1–25. https://doi.org/10.1080/00437956.2006.11432557

Conti-Ramsden, G., Ullman, M. T., & Lum, J. A. G. (2015). The relation between receptive grammar and procedural, declarative, and working memory in specific language impairment. *Frontiers in Psychology*, *6*, 1090. https://doi.org/

10.3389/fpsyg.2015.01090

Conway, A., Jarrold, C., & Kane, M. J. (2007). *Variation in working memory*. Oxford: Oxford University Press.

Conway, A., Kane, M. J., Bunting, M. F., Hambrick, D. Z., Wilhelm, O., & Engle, R. W. (2005). Working memory span tasks: A methodological review and user's guide. *Psychonomic Bulletin & Review*, *12*(5), 769–786. https://doi.org/10.3758/BF03196772

Cornell, A. (1985). Realistic goals in teaching and learning phrasal verbs. *International Review of Applied Linguistics in Language Teaching*, *21*(1-4), 269–280. https://doi.org/10.1515/iral.1985.23.1-4.269

Corno, L., & Snow, R. E. (1986). Adapting teaching to individual differences among learners. In M. C. Wittrock (Ed.), *Handbook of research on teaching (3rd. ed.)*. New York: Macmillan.

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

Cowan, N. (1988). Evolving conceptions of memory storage, selective attention, and their mutual constraints within the human information-processing system. *Psychological Bulletin*, *104*(2), 163. https://doi.org/10.1037/0033-2909.104.2.163

Cowan, N. (2010). The magical mystery four: How is working memory capacity limited, and why? *Current directions in psychological science*, *19*(1), 51–57. https://doi.org/10.1177/0963721409359277

Cowan, N. (2015). Second language use, theories of working memory and the vennian mind. In Z. Wen, M. Borges Mota, & A. McNeill (Eds.), *Working memory in second language acquisition and processing* (pp. 29–40). Bristol: Multilingual Matters.

Cowan, N. (2016). Working memory maturation: Can we get at the essence of cognitive growth? *Perspectives on Psychological Science*, *11*(2), 239-264. https://doi.org/10.1177/1745691615621279

Cowan, N. (2017). The many faces of working memory and short-term storage. *Psycho-*

nomic Bulletin & Review, *24*(4), 1158–1170. https://doi.org/10.3758/s13423-016-1191-6

Craik, F. I., & Tulving, E. (1975). Depth of processing and the retention of words in episodic memory. *Journal of Experimental Psychology: General*, *104*(3), 268–294. https://doi.org/10.1037/0096-3445.104.3.268

Creswell, J. W. (2015). *Educational research: Planning, conducting, and evaluating quantitative and qualitative research* (4th ed.). Boston: Pearson Education.

Cronbach, L. J. (1957). The two disciplines of scientific psychology. *American Psychologist*, *12*(11), 671. https://doi.org/10.1037/h0043943

Cronbach, L. J., & Snow, R. E. (1977). *Aptitudes and instructional methods: A handbook for research on interactions*. England: Irvington.

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

Cunnings, I. (2012). An overview of mixed-effects statistical models for second language researchers. *Second Language Research*, *28*(3), 369–382. https://doi.org/10.1177/0267658312443651

Cunnings, I., & Finlayson, I. (2015). Mixed effects modelling and longitudinal data analysis. In L. Plonsky (Ed.), *Advancing quantitative methods in second language research* (pp. 159–181). Abingdon: Routledge.

Dagut, M., & Laufer, B. (1985). Avoidance of phrasal verbs—A case for contrastive analysis. *Studies in Second Language Acquisition*, *7*(1), 73–79. https://doi.org/10.1017/S0272263100005167

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

Daskalovska, N. (2011). The impact of reading on three aspects of word knowledge: Spelling, meaning and collocation. *Procedia - Social and Behavioral Sciences*, *15*, 2334 - 2341. https://doi.org/10.1016/j.sbspro.2011.04.103

Davies, M. (2009). The 385+ million word Corpus of Contemporary American English (1990–2008+): Design, architecture, and linguistic insights. *International Journal of Corpus Linguistics*, *14*(2), 159–190. https://doi.org/10.1075/ijcl.14.2.02dav

de Graaff, R., & Housen, A. (2009). Investigating the effects and effectiveness of L2 instruction. In M. H. Long & C. J. Doughty (Eds.), *The handbook of language teaching* (pp. 726–755). Oxford: Blackwell.

Dehé, N. (2002). *Particle verbs in English: Syntax, information structure and intonation*. Amsterdam: John Benjamins Publishing.

DeKeyser, R. M. (1995). Learning second language grammar rules: An experiment with a miniature linguistic system. *Studies in Second Language Acquisition*, *17*(3), 379–410. https://doi.org/10.1017/S027226310001425X

DeKeyser, R. M. (1997). Beyond explicit rule learning: Automatizing second language morphosyntax. *Studies in Second Language Acquisition*, *19*(2), 195–221.

DeKeyser, R. M. (2012a). Age effects in second language learning. In *The Routledge handbook of second language acquisition* (pp. 442–460). New York: Routledge.

DeKeyser, R. M. (2012b). Interactions between individual differences, treatments, and structures in SLA. *Language Learning*, *62*, 189–200. https://doi.org/10.1111/j.1467-9922.2012.00712.x

DeKeyser, R. M. (2016). Of moving targets and chameleons: Why the concept of difficulty is so hard to pin down. *Studies in Second Language Acquisition*, *38*(2), 353–363. https://doi.org/10.1017/S0272263116000024

DeKeyser, R. M., & Koeth, J. (2011). Cognitive aptitudes for second language learning. In E. Hinkel (Ed.), *Handbook of research in second language teaching and learning, Volume II* (pp. 395–406). London: Routledge.

Della Putta, P. (2016). The effects of textual enhancement on the acquisition of two nonparallel grammatical features by Spanish-speaking learners of Italian. *Studies in Second Language Acquisition*, *38*(2), 217–238. https://doi.org/10.1017/S0272263116000073

Denhovska, N., & Serratrice, L. (2017). Incidental learning of gender agreement in L2. *Journal of Psycholinguistic Research*, *46*(5), 1187–1211. https://doi.org/10.1007/s10936-017-9487-x

Denhovska, N., Serratrice, L., & Payne, J. (2018). Frequency and working memory effects in incidental learning of a complex agreement pattern. *Lingua*. https://doi.org/10.1016/j.lingua.2018.02.009

Deshors, S. C. (2016). Inside phrasal verb constructions: A co-varying collexeme analysis of verb-particle combinations in EFL and their semantic associations. *International Journal of Learner Corpus Research*, *2*(1), 1–30.

Dewaele, J.-M. (2017). Psychological dimensions and foreign language anxiety. In S. Loewen & M. Sato (Eds.), *The Routledge handbook of instructed second language acquisition* (pp. 433–467). New York: Routledge.

Dewaele, J.-M., & Al-Saraj, T. M. (2015). Foreign language classroom anxiety of Arab learners of English: The effect of personality, linguistic and sociobiographical variables. *Studies in Second Language Learning and Teaching*, *5*(2), 205–228. https://doi.org/10.14746/ssllt.2015.5.2.2

Dickinson, M., Brew, C., & Meurers, D. (2013). *Language and computers*. Oxford: Wiley-Blackwell.

Diessel, H., & Tomasello, M. (2005). Particle placement in early child language: A multifactorial analysis. *Corpus Linguistics and Linguistic Theory*, *1*(1), 89–112. https://doi.org/10.1515/cllt.2005.1.1.89

Dörnyei, Z. (2003). Individual differences in foreign language learning: Effects of aptitude, intelligence, and motivation. *Studies in Second Language Acquisition*, 584–585. https://doi.org/10.1017/S0272263103220258

Dörnyei, Z. (2005). *The psychology of the language learner: Individual differences in second language acquisition*. New York: Lawrence Erlbaum Associates, Inc.

Dörnyei, Z., & Katona, L. (1992). Validation of the C-test amongst Hungarian EFL learners. *Language Testing*, *9*(2), 187–206. https://doi.org/10.1177/026553229200900206

Dörnyei, Z., & Ushioda, E. (2012). Motivation. In *The Routledge handbook of second language acquisition* (pp. 396–409). New York: Routledge.

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

Doughty, C. J. (2003). Instructed SLA: Constraints, compensation, and enhancement. In C. J. Doughty & M. H. Long (Eds.), *The handbook of second language acquisition* (pp. 206–257). New York: Blackwell.

Doughty, C. J., & Williams, J. (1998). Pedagogical choices in focus on form. In C. Doughty & J. Williams (Eds.), *Focus on form in classroom second language acquisition* (pp. 197–261). Cambridge: Cambridge University Press.

Downing, A., & Locke, P. (2006). *English grammar: A university course*. London/New York: Routledge.

Eckes, T., & Grotjahn, R. (2006). A closer look at the construct validity of C-tests. *Language Testing*, *23*(3), 290–325. https://doi.org/10.1191/0265532206lt330oa

Eichenbaum, H. (2004). Hippocampus: Cognitive processes and neural representations that underlie declarative memory. *Neuron*, *44*(1), 109–120. https://doi.org/10.1016/j.neuron.2004.08.028

Eichenbaum, H. (2011). *The cognitive neuroscience of memory: An introduction*. New York: Oxford University Press.

El-Dakhs, D. A. S. (2016). The lexical knowledge and avoidance of phrasal verbs: The case of Egyptian learners of English. *International Journal of Applied Linguistics and English Literature*, *5*(1), 132–144. https://doi.org/10.7575/aiac.ijalel.v.5n.1p.132

Elliott, G. C. (1998). Interpreting higher order interactions in log-linear analysis. *Psychological Bulletin*, *103*(1), 121–130. https://doi.org/10.1037/0033-2909.103.1.121

Ellis, R. (1994). *The study of second language acquisition*. Cambridge: Oxford University Press.

Ellis, R. (2001). *Form-focused instruction and second language learning: Language learning*. Malden, MA: Blackwell.

Ellis, R. (2008). Explicit form-focused instruction and second language acquisition. In B. Spolsky & F. M. Hult (Eds.), *The handbook of educational linguistics* (pp. 437–455). New York: Blackwell.

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

Ellis, R. (2015a). The importance of focus on form in communicative language teaching. *Eurasian Journal of Applied Linguistics*, *1*(2), 1–15. Retrieved from http://www.ejal.eu/index.php/ejal/article/view/50

Ellis, R. (2015b). *Understanding second language acquisition* (2nd ed.). Oxford: Oxford Universty Press.

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

Ellis, R. (2017). Task-based language teaching. In S. Loewen & M. Sato (Eds.), *The Routledge handbook of instructed second language acquisition* (pp. 108–125). New York: Routledge.

Ellis, R. (in press). Foreword. In Z. Wen, P. Skehan, A. Biedroń, S. Li, & R. Sparks (Eds.), *Language aptitude: Multiple perspectives and emerging trends.* Routledge.

Engle, R. W. (2002). Working memory capacity as executive attention. *Current Directions in Psychological Science*, *11*(1), 19–23. https://doi.org/10.1111/1467-8721.00160

Engle, R. W., Kane, M. J., & Tuholski, S. W. (1999). Individual differences in working memory capacity and what they tell us about controlled attention, general fluid intelligence, and functions of the prefrontal cortex. In A. Miyake & P. Shah (Eds.), *Models of working memory: Mechanisms of active maintenance and executive control* (pp. 102–134). Cambridge University Press.

Erçetin, G., & Alptekin, C. (2013). The explicit/implicit knowledge distinction and working memory: Implications for second-language reading comprehension. *Applied Psycholinguistics*, *34*(4), 727–753. https://doi.org/10.1017/S0142716411000932

Erlam, R. (2005). Language aptitude and its relationship to instructional effectiveness in second language acquisition. *Language Teaching Research*, *9*(2), 147–171. https://doi.org/10.1191/1362168805lr161oa

Erteschik-Shir, N. (1979). Discourse constraints on dative movement. In T. Givón (Ed.), *Discourse and syntax: Syntax and semantics* (Vol. 12, pp. 441–467). New York: Academic Press.

Ettlinger, M., Bradlow, A. R., & Wong, P. C. (2014). Variability in the learning of complex morphophonology. *Applied Psycholinguistics*, *35*(4), 807–831. https://doi.org/10.1017/S0142716412000586

Evison, J. (2010). What are the basis of analysing a corpus. In A. O'Keeffe & M. McCarthy (Eds.), *The Routledge handbook of corpus linguistics* (pp. 122–135). New York: Routledge.

Fadanelli, S. B. (2012). A corpus study on Brazilian learners usage of English phrasal verbs. *Estudos Anglo-Americanos*, *37*, 158–176.

Fang, T. (2016). Effects of textual enhancement on English as a foreign language learners' anaphor resolution performance and reading comprehension in Taiwan. *Journal of Research in Reading*, *39*(3), 347–365. https://doi.org/10.1111/1467-9817.12028

Faretta-Stutenberg, M., & 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. https://doi.org/10.1177/0267658316684903

Farsani, H. M., Moinzadeh, A., & Tavakoli, M. (2012). Mnemonic effectiveness of CL-motivated picture-elucidation tasks in foreign learners' acquisition of English phrasal verbs. *Theory and Practice in Language Studies*, *2*(3), 498–509.

https://doi.org/10.4304/tpls.2.3.498-509

Fecteau, M. L. (1999). First-and second-language reading comprehension of literary texts. *The Modern Language Journal*, *83*(4), 475–493. https://doi.org/10.1111/0026-7902.00036

Fletcher, P., Shallice, T., Frith, C., Frackowiak, R., & Dolan, R. (1998). The functional roles of prefrontal cortex in episodic memory. ii. retrieval. *Brain*, *121*(7), 1249–1256. https://doi.org/10.1093/brain/121.7.1249

Fontanini, I., & Tomitch, L. M. B. (2009). Working memory capacity and L2 university students comprehension of linear texts and hypertexts. *International Journal of English Studies*, *9*(2). Retrieved from http://revistas.um.es/ijes/article/view/90721

Fraser, B. (1974). *The verb-particle combination in English.* Tokyo: Taisbukan.

Fuchs, L. S., Schumacher, R. F., Sterba, S. K., Long, J., Namkung, J., Malone, A., … others (2014). Does working memory moderate the effects of fraction intervention? an aptitude–treatment interaction. *Journal of Educational Psychology*, *106*(2), 499. https://doi.org/10.1037/a0034341

Gablasova, D., Brezina, V., & McEnery, T. (2017). Collocations in corpus-based language learning research: Identifying, comparing, and interpreting the evidence. *Language Learning*, *67*(S1), 155–179. https://doi.org/10.1111/lang.12225

Gamper, J., & Knapp, J. (2002). A review of intelligent CALL systems. *Computer Assisted Language Learning*, *15*(4), 329–342. https://doi.org/10.1076/call.15.4.329.8270

Ganji, M. (2011). The best way to teach phrasal verbs: Translation, sentential contextualization or metaphorical conceptualization. *Theory and Practice in Language Studies*, *1*(11), 1497–1506. https://doi.org/10.4304/tpls.1.11.1497-1506

Gardner, D., & Davies, M. (2007). Pointing out frequent phrasal verbs: A corpus-based analysis. *TESOL Quarterly*, *41*(2), 339–359. https://doi.org/10.1002/j.1545-7249.2007.tb00062.x

Garnier, M. (2016). *English phrasal verbs: Usage, knowledge, acquisition* (Doctoral dissertation). University of Nottingham, Nottingham, United Kingdom.

Garnier, M., & Schmitt, N. (2015). The PHaVE List: A pedagogical list of phrasal verbs and their most frequent meaning senses. *Language Teaching Research*, *19*(6), 645–666. https://doi.org/10.1177/1362168814559798

Garnier, M., & Schmitt, N. (2016). Picking up polysemous phrasal verbs: How many do learners know and what facilitates this knowledge? *System*, *59*, 29–44. https://doi.org/10.1016/j.system.2016.04.004

Gavioli, L. (2005). *Exploring corpora for ESP learning*. Amsterdam: John Benjamins Publishing.

Gelman, A. (2018). The failure of null hypothesis significance testing when studying incremental changes, and what to do about it. *Personality and Social Psychology Bulletin*, *44*(1), 16–23. https://doi.org/10.1177/0146167217729162

Gharedaghi, M., & Touran, A. (2016). A comparative study of the effect of explicit vs. implicit instruction on learning phrasal verbs by Iranian high school students. *Modern Journal of Language Teaching Methods*, *6*(4), 401–413.

Gilquin, G. (2015). The use of phrasal verbs by French-speaking EFL learners. A constructional and collostructional corpus-based approach. *Corpus Linguistics and Linguistic Theory*, *11*(1), 51–88. https://doi.org/10.1515/cllt-2014-0005

Gonthier, C., Aubry, A., & Bourdin, B. (2017). Measuring working memory capacity in children using adaptive tasks: Example validation of an adaptive complex span. *Behavior Research Methods*. https://doi.org/10.3758/s13428-017-0916-4

González Fernández, B., & Schmitt, N. (2015). How much collocation knowledge do L2 learners have? *International Journal of Applied Linguistics*, *166*(1), 94–126.

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

Goo, J. (2012). Corrective feedback and working memory capacity in interaction-driven L2 learning. *Studies in Second Language Acquisition*, *34*(3), 445–474. https://doi.org/10.1017/S0272263112000149

Goo, J., Granena, G., Yilmaz, Y., & 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* (pp. 443–482). Amsterdam/Philadelphia: John Benjamins Publishing.

Grabe, W. (2009). *Reading in a second language: Moving from theory to practice*. New York: Cambridge University Press.

Greenberg, J. A., Burke, J. F., Haque, R., Kahana, M. J., & Zaghloul, K. A. (2015). Decreases in theta and increases in high frequency activity underlie associative memory encoding. *Neuroimage*, *114*, 257–263. https://doi.org/10.1016/j.neuroimage.2015.03.077

Greenslade, T., Bouden, L., & Sanz, C. (1999). Attending to form and content in processing L2 reading texts. *Spanish Applied Linguistics*, *3*(1), 65–90.

Gregersen, T., & MacIntyre, P. D. (2013). *Capitalizing on language learners' individuality: From premise to practice*. Bristol: Multilingual Matters.

Grey, S., Williams, J. N., & Rebuschat, P. (2015). Individual differences in incidental language learning: Phonological working memory, learning styles, and personality. *Learning and Individual Differences*, *38*, 44–53. https://doi.org/10.1016/j.lindif.2015.01.019

Gries, S. T. (2011). Acquiring particle placement in English: A corpus-based perspective. In P. G. Medina (Ed.), *Morphosyntactic alternations in English: Functional and cognitive perspectives.* London/Oakville: CT: Equinox.

Groot, P. J. M. (2000). Computer assisted second language vocabulary acquisition. *Language Learning & Technology*, *4*(1), 60–81. Retrieved from http://llt.msu.edu/vol4num1/groot/default.html

Grundy, J. G., & Timmer, K. (2017). Bilingualism and working memory capacity: A comprehensive meta-analysis. *Second Language Research*, *33*(3), 325–340.

https://doi.org/10.1177/0267658316678286

Guo, X., & Yang, Y. (2018). Effects of corrective feedback on EFL learners' acquisition of third-person singular form and the mediating role of cognitive style. *Journal of Psycholinguistic Research*, 1–18. https://doi.org/10.1007/s10936-018-9566-7

Gwaltney, C. J., Shields, A. L., & Shiffman, S. (2008). Equivalence of electronic and paper-and-pencil administration of patient-reported outcome measures: A meta-analytic review. *Value in Health*, *11*(2), 322–333. https://doi.org/10.1111/j.1524-4733.2007.00231.x

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

Hamrick, P., Lum, J. A. G., & Ullman, M. T. (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. https://doi.org/10.1073/pnas.1713975115

Han, Z., Park, E. S., & Combs, C. (2008). Textual enhancement of input: Issues and possibilities. *Applied Linguistics*, *29*(4), 597–618. https://doi.org/10.1093/applin/amn010

Harding, L., Alderson, J. C., & Brunfaut, T. (2015). Diagnostic assessment of reading and listening in a second or foreign language: Elaborating on diagnostic principles. *Language Testing*, *32*(3), 317–336. https://doi.org/10.1177/0265532214564505

Hare, A. (2010). *Incidental phrasal verb acquisition through second language reading* (mathesis). Concordia University, Canada.

Harrington, M., & Sawyer, M. (1992). L2 working memory capacity and L2 reading skill. *Studies in Second Language Acquisition*, *14*(1), 25–38. https://doi.org/10.1017/S0272263100010457

Harsch, C., & Hartig, J. (2016). Comparing C-tests and Yes/No vocabulary size tests as predictors of receptive language skills. *Language Testing*, *33*(4), 555–575. https://doi.org/10.1177/0265532215594642

Hassel, M. (2001). *Internet as corpus: Automatic construction of a Swedish news corpus* (Tech. Rep.). KTH, Numerical Analysis and Computer Science, NADA. Retrieved from http://urn.kb.se/resolve?urn=urn:nbn:se:kth:diva-14075

Hassel, M. (2007). *Resource lean and portable automatic text tummarization* (Doctoral dissertation, KTH, Numerical Analysis and Computer Science, NADA, Stockholm, Sweden). Retrieved from http://urn.kb.se/resolve?urn=urn:nbn:se:kth:diva-4414

Heift, T. (2005). Inspectable learner reports for web-based language learning. *ReCALL*, *17*(1), 32–46. https://doi.org/10.1017/S0958344005000418

Heift, T. (2010). Developing an intelligent language tutor. *CALICO Journal*, *27*(3), 443–459.

Heift, T. (2013). Intelligent computer-assisted language learning. In *The encyclopedia of applied linguistics.* https://doi.org/10.1002/9781405198431.wbeal0548

Heift, T. (2016). Web delivery of adaptive and interactive language tutoring: Revisited. *International Journal of Artificial Intelligence in Education*, *26*(1), 489–503. https://doi.org/10.1007/s40593-015-0061-0

Heift, T. (2017). History and key developments in intelligent computer-assisted language learning (ICALL). In S. L. Thorne & S. May (Eds.), *Language, education and technology* (pp. 289–300). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-319-02237-6_23

Heift, T., & Rimrott, A. (2012). Task-related variation in computer-assisted language learning. *The Modern Language Journal*, *96*(4), 525–543. https://doi.org/10.1111/j.1540-4781.2012.01392.x

Heift, T., & Schulze, M. (2007). *Errors and intelligence in computer-assisted language learning: Parsers and pedagogues*. London: Routledge.

Heift, T., & Vyatkina, N. (2017). Technologies for teaching and learning L2 grammar. In C. A. Chapelle & S. Sauro (Eds.), *The handbook of technology and second language teaching and learning* (pp. 26–44). Oxford: Wiley-Blackwell.

Hicks, K. L., Foster, J. L., & Engle, R. W. (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. https://doi.org/10.1016/j.jarmac.2016.07.010

Hong, L., & MacWhinney, B. (2011). Semantic competitor priming within and across languages: The interplay of vocabulary knowledge, learning experience and working memory capacity. *Bilingualism: Language and Cognition*, *14*(4), 433–443. https://doi.org/10.1017/S1366728910000507

Honnibal, M., & Johnson, M. (2015). An improved non-monotonic transition system for dependency parsing. In *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing* (pp. 1373–1378). Lisbon, Portugal: Association for Computational Linguistics.

Hopman, E. W. M., & MacDonald, M. C. (2018). Production practice during language learning improves comprehension. *Psychological Science*. https://doi.org/10.1177/0956797618754486

Horzum, M. B., İsmail Önder, & Şenol Beşoluk. (2014). Chronotype and academic achievement among online learning students. *Learning and Individual Differences*, *30*, 106–111. https://doi.org/10.1016/j.lindif.2013.10.017

Housen, A., & Pierrard, M. (2005). Investigating instructed second language acquisition. In A. Housen & M. Pierrard (Eds.), *Investigations in instructed second language acquisition* (pp. 1–27). Berlin: Mouton de Gruyter.

Howarth, P. (1998). Phraseology and second language proficiency. *Applied Linguistics*, *19*(1), 24–44. https://doi.org/10.1093/applin/19.1.24

Huang, L.-l., & Lin, C.-c. (2014). Three approaches to glossing and their effects on vocabulary learning. *System*, *44*, 127–136. https://doi.org/10.1016/j.system.2014.03.006

Huang, L.-S. (2017). Has corpus-based instruction reached a tipping point? Practical applications and pointers for teachers. *TESOL Journal*, *8*(2), 295–313. https://doi.org/10.1002/tesj.271

Huebner, T. (1983). Linguistic systems and linguistic change in an interlanguage. *Studies in Second Language Acquisition*, *6*(1), 33–53. https://doi.org/10.1017/S0272263100000280

Hulstijn, J. H. (2001). Intentional and incidental second language vocabulary learning: A reappraisal of elaboration, rehearsal and automaticity. In P. Robinson (Ed.), *Cognition and second language instruction* (pp. 258–286). Cambrigde University Press.

Hulstijn, J. H. (2003). Incidental and intentional learning. In C. J. Doughty & M. H. Long (Eds.), *The handbook of second language acquisition* (pp. 349–381). Oxford: Blackwell.

Hulstijn, J. H. (2005). Theoretical and empirical issues in the study of implicit and explicit second-language learning: Introduction. *Studies in Second Language Acquisition*, *27*(2), 129–140. https://doi.org/10.1017/S0272263105050084

Hulstijn, J. H., & Laufer, B. (2001). Some empirical evidence for the involvement load hypothesis in vocabulary acquisition. *Language Learning*, *51*(3), 539–558. https://doi.org/10.1111/0023-8333.00164

Hulstijn, J. H., & Marchena, E. (1989). Avoidance: Grammatical or semantic causes? *Studies in Second Language Acquisition*, *11*(3), 241–255. https://doi.org/10.1017/S0272263100008123

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

Hutcheson, G. (2011). Categorical explanatory variables. *Journal of Modelling in Management*, *6*(2). https://doi.org/10.1108/jm2.2011.29706baa.002

Hwu, F., Pan, W., & Sun, S. (2014). Aptitude-treatment interaction effects on explicit rule learning: A latent growth curve analysis. *Language Teaching Research*, *18*(3), 294–319. https://doi.org/10.1177/1362168813510381

Hwu, F., & Sun, S. (2012). The aptitude-treatment interaction effects on the learning of grammar rules. *System*, *40*(4), 505–521. https://doi.org/10.1016/j.system.2012.10.009

Indrarathne, B., & Kormos, J. (2017). Attentional processing of input in explicit and implicit conditions: An eye-tracking study. *Studies in Second Language Acquisition*, *39*(3), 401–430. https://doi.org/10.1017/S027226311600019X

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

Indrarathne, B., Ratajczak, M., & Kormos, J. (2018). Modelling changes in the cognitive processing of grammar in implicit and explicit learning conditions: Insights from an eye-tracking study. *Language Learning*. https://doi.org/10.1111/lang.12290

Ingvalson, E. M., Nowicki, C., Zong, A., & Wong, P. C. M. (2017). Non-native speech learning in older adults. *Frontiers in Psychology*, *8*, 148. https://doi.org/10.3389/fpsyg.2017.00148

Issa, B., Morgan-Short, K., Villegas, B., & Raney, G. (2015). An eye-tracking study on the role of attention and its relationship with motivation. In Roberts, K. McManus, N. Vanek, & D. Trenkic (Eds.), *EUROSLA Yearbook 15* (pp. 114–142). Amsterdam: John Benjamins.

Izumi, S. (2002). Output, input enhancement, and the noticing hypothesis: An experimental study on ESL relativization. *Studies in Second Language Acquisition*, *24*(4), 541–577. https://doi.org/10.1017/S0272263102004023

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

Jahan, A., & Kormos, J. (2015). The impact of textual enhancement on EFL learners' grammatical awareness of future plans and intentions. *International Jour-*

nal of Applied Linguistics, *25*(1), 46–66. https://doi.org/10.1111/ijal.12049

Jeon, E. H., & Yamashita, J. (2014). L2 reading comprehension and its correlates: A meta-analysis. *Language Learning*, *64*(1), 160–212. https://doi.org/10.1111/lang.12034

Joh, J., & Plakans, L. (2017). Working memory in L2 reading comprehension: The influence of prior knowledge. *System*, *70*, 107–120. https://doi.org/10.1016/j.system.2017.07.007

Johns, T. (1990). From printout to handout: Grammar and vocabulary teaching in the context of data-driven learning. *CALL Austria*, *10*, 14–34.

Jonassen, D. H., & Grabowski, B. L. (1993). *Handbook of individual differences, learning, and instruction*. Hillsdale, New Jersey: Lawrence Erlbaum Associates, Publishers.

Jordano, M. L., & Touron, D. R. (2017). Stereotype threat as a trigger of mind-wandering in older adults. *Psychology and Aging*, *32*(3), 307 - 313. https://doi.org/10.1037/pag0000167

Jourdenais, R. (1998). *The effects of textual enhancement on the acquisition of the Spanish preterit and imperfect* (Doctoral dissertation). Georgetown University, Washington, DC.

Jourdenais, R., Ota, M., Stauffer, S., Boyson, B., & Doughty, C. J. (1995). Does textual enhancement promote noticing? A think-aloud protocol analysis. In R. W. Schmidt (Ed.), *Attention and awareness in foreign language learning* (pp. 183–216). Honolulu: University of Hawai‘i, Second Language Teaching & Curriculum Center.

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

Jung, J. (2016). Effects of glosses on learning of L2 grammar and vocabulary. *Language Teaching Research*, *20*(1), 92-112. https://doi.org/10.1177/1362168815571151

Jung, J. (2017). *Effects of task complexity, glossing and working memory on L2 reading and L2 learning* (Doctoral dissertation). University College London, London, United Kingdom.

Jung, J. (2018). Effects of task complexity and working memory capacity on L2 reading comprehension. *System*, *74*, 21–37. https://doi.org/10.1016/j.system.2018.02.005

Jurafsky, D., & Martin, J. H. (2009). *Speech and language processing: An introduction to natural language processing, computational linguistics, and speech recognition*. Upper SaddleRiver, NJ: Prentice Hall.

Jylkkä, J., Soveri, A., Wahlström, J., Lehtonen, M., Rodríguez-Fornells, A., & Laine, M. (2017). Relationship between language switching experience and executive functions in bilinguals: An Internet-based study. *Journal of Cognitive Psychology*, *29*(4), 404–419. https://doi.org/10.1080/20445911.2017.1282489

Kachinske, I. (2016). *The role of rules, examples and individual differences in the acquisition of declarative and procedural second language knowledge* (Doctoral dissertation). University of Maryland, College Park, Maryland, United States.

Kadia, K. (1998). The effect of formal instruction on monitored and on spontaneous naturalistic interlanguage performance: A case study. *TESOL Quarterly*, *22*(3), 509–515. https://doi.org/10.2307/3587295

Kamarudin, R. (2013). *A study on the use of phrasal verbs by Malaysian learners of English* (Doctoral dissertation). University of Birmingham, Birmingham, United Kingdom.

Kane, M. J., Conway, A. R., Hambrick, D. Z., & Engle, R. W. (2007). Variation in working memory capacity as variation in executive attention and control. In A. Conway, C. Jarrold, & M. J. Kane (Eds.), *Variation in working memory* (pp. 21–48). Oxford: Oxford University Press.

Kane, M. J., Hambrick, D. Z., Tuholski, S. W., Wilhelm, O., Payne, T. W., & Engle, R. W. (2004). The generality of working memory capacity: A latent-variable approach to verbal and visuospatial memory span and reasoning. *Journal of Experimental Psychology: General*, *133*(2), 189. https://doi.org/10.1037/0096-3445.133.2.189

Kang, E. Y., Sok, S., & Han, Z. (2018). Thirty-five years of ISLA on form-focused instruction: A meta-analysis. *Language Teaching Research*. https://doi.org/10.1177/1362168818776671

Kartal, G., & Uner, S. (2017). The effects of conceptual metaphors on the acquisition of phrasal verbs by Turkish EFL learners. *European Journal of Foreign Language Teaching*, *2*(2). Retrieved from http://www.oapub.org/edu/index.php/ejfl/article/view/685

Ke, Y. (2014). *The phraseology of phrasal verbs in English: A corpus study of the language of Chinese learners and native English writers* (Doctoral dissertation, University of Birmingham, Birmingham, United Kingdom). Retrieved from http://ethos.bl.uk/OrderDetails.do?uin=uk.bl.ethos.629721

Ke, Y. (2016). A bi-axis model for profiling English phrasal verbs for pedagogic purposes. *TESOL Quarterly*. https://doi.org/10.1002/tesq.345

Kempe, V., Brooks, P. J., & Christman, S. D. (2009). Inconsistent handedness is linked to more successful foreign language vocabulary learning. *Psychonomic Bulletin & Review*, *16*(3), 480–485. https://doi.org/10.3758/PBR.16.3.480

Kempe, V., Brooks, P. J., & Kharkhurin, A. (2010). Cognitive predictors of generalization of russian grammatical gender categories. *Language Learning*, *60*(1), 127–153.

Khatib, M., & Ghannadi, M. (2011). Interventionist (explicit and implicit) versus non-interventionist (incidental) learning of phrasal verbs by Iranian EFL learners. *Journal of Language Teaching & Research*, *2*(3), 537–546. https://doi.org/10.4304/jltr.2.3.537-546

Khezrlou, S. (2018). Form-focused instruction in CALL: What do learners think? *RELC Journal*. https://doi.org/10.1177/0033688217738820

Khezrlou, S., Ellis, R., & Sadeghi, K. (2017). Effects of computer-assisted glosses on EFL learners' vocabulary acquisition and reading comprehension in three learning conditions. *System*, *65*, 104–116. https://doi.org/10.1016/j.system.2017.01.009

Khodareza, M. R., & Shabani, G. (2016). Comparing the effect of one-way versus two-way tasks on Iranian intermediate EFL learners' phrasal verb learning. *Journal of Studies in Education*, *6*(1), 129–146. https://doi.org/10.5296/jse.v6i1.8799

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

Kim, J., & Nassaji, H. (2017). Incidental focus on form and the role of learner extraversion. *Language Teaching Research*. https://doi.org/10.1177/1362168817699642

Kim, K., & Clariana, R. B. (2017). Text signals influence second language expository text comprehension: Knowledge structure analysis. *Educational Technology Research and Development*, *65*(4), 909–930. https://doi.org/10.1007/s11423-016-9494-x

Kim, Y. (2006). Effects of input elaboration on vocabulary acquisition through reading by Korean learners of English as a foreign language. *TESOL Quarterly*, *40*(2), 341–373. https://doi.org/10.2307/40264526

King, K. A., & Mackey, A. (2016). Research methodology in second language studies: Trends, concerns, and new directions. *The Modern Language Journal*, *100*(S1), 209–227. https://doi.org/10.1111/modl.12309

Kintsch, W., & Van Dijk, T. A. (1978). Toward a model of text comprehension and production. *Psychological review*, *85*(5), 363. https://doi.org/10.1037/0033-295X.85.5.363

Klaus, J., & Schriefers, H. (2016). *Measuring verbal working memory capacity: A reading span task for laboratory and web-based use.* Retrieved from osf.io/geuwk

Klein, W. (1986). *Second language acquisition.* Cambridge: Cambridge University

Press.

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. https://doi.org/10.1177/026553228500200108

Knowlton, B. J., Siegel, A. L., & Moody, T. D. (2017). Procedural learning in humans. In J. H. Byrne (Ed.), *Learning and memory: A comprehensive reference (second edition)* (pp. 295–312). Oxford: Academic Press. https://doi.org/10.1016/B978-0-12-809324-5.21085-7

Koda, K. (2012). How to do research on second language reading. In A. Mackey & S. M. Gass (Eds.), *Research methods in second language acquisition: A practical guide* (pp. 158–179). Malden, MA: Blackwell.

Kormos, J. (2013). New conceptualizations of language aptitudein second language attainment. In G. Granena & M. H. Long (Eds.), *Sensitive periods, language aptitude and ultimate L2 attainment* (pp. 131–152). Amsterdam: John Benjamins.

Kövecses, Z., & Szabcó, P. (1996). Idioms: A view from cognitive semantics. *Applied Linguistics*, *17*(3), 326–355. https://doi.org/10.1093/applin/17.3.326

Krantz, J. H., Ballard, J., & Scher, J. (1997). Comparing the results of laboratory and World-Wide Web samples on the determinants of female attractiveness. *Behavior Research Methods, Instruments, & Computers*, *29*(2), 264–269. https://doi.org/10.3758/BF03204824

Krantz, J. H., & Reips, U.-D. (2017). The state of web-based research: A survey and call for inclusion in curricula. *Behavior Research Methods*, *49*(5), 1621–1629. https://doi.org/10.3758/s13428-017-0882-x

Krashen, S. D. (1981). *Second language acquisition and second language learning*. Oxford: Pergamon.

Krashen, S. D. (1985). *The input hypothesis: Issues and implications*. London: Longman.

Krashen, S. D., & Terrell, T. D. (1983). *The natural approach: Language acquisition in* *the classroom*. New York: Pergamon Press.

Kreuz, R. J., & Caucci, G. M. (2007). Salience as a concept in cognitive psychology. In C. Gascoigne (Ed.), *Assessing the impact of input enhancement in second language education: Evolution in theory, research and practice* (pp. 19–35). Stillwater, OK: New Forums Press.

Kubota, M. (1997). Instructional effects of positive and negative evidence on prepositional/phrasal verbs. *Institute for Research in Language Teaching. Bulletin*, *11*(1–39).

Kubota, S. (2000). *Input enhancement in classroom second language acquisition of Japanese* (Doctoral dissertation). Purdue University, West Lafayette, IN.

Kyllonen, P. C., & Christal, R. E. (1990). Reasoning ability is (little more than) working-memory capacity?! *Intelligence*, *14*(4), 389–433. https://doi.org/10.1016/S0160-2896(05)80012-1

LaBrozzi, R. M. (2016). The effects of textual enhancement type on L2 form recognition and reading comprehension in Spanish. *Language Teaching Research*, *20*(1), 75–91. https://doi.org/10.1177/1362168814561903

Lado, B. (2017). Aptitude and pedagogical conditions in the early development of a nonprimary language. *Applied Psycholinguistics*, *38*(3), 679–701. https://doi.org/10.1017/S0142716416000394

Lado, B., & Sanz, C. (in press). Feedback, aptitude and multilingualism. In H. Nassaji & E. Kartchava (Eds.), *The Cambridge handbook of corrective feedback in language learning and teaching*. Cambridge: Cambridge University Press.

Lakoff, G., & Johnson, M. (1980). *Metaphors we live by*. Chicago, IL: University of Chicago Press.

Larsen-Freeman, D. (1991). Teaching grammar. In M. Celce-Murcia (Ed.), *Teaching English as a second or foreign language* (2nd ed., pp. 279–296). New York: Newbury House.

Larsen-Freeman, D. (2009). Teaching and testing grammar. In M. H. Long & C. J. Doughty (Eds.), *The handbook of language teaching* (pp. 518–542). Oxford:

Blackwell.

Larsen-Freeman, D. (2018). Looking ahead: Future directions in, and future research into, second language acquisition. *Foreign Language Annals*. https://doi.org/10.1111/flan.12314

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

Laufer, B. (2005). Focus on form in second language vocabulary learning. In *EUROSLA Yearbook 5* (pp. 223–250).

Laufer, B. (2012). Second language word difficulty. In C. A. Chapelle (Ed.), *The encyclopedia of applied linguistics* (pp. 5151–5156). Oxford: Wiley-Blackwell.

Laufer, B., & Eliasson, S. (1993). What causes avoidance in L2 learning: L1-L2 difference, L1-L2 similarity, or L2 complexity? *Studies in Second Language Acquisition*, *15*(1), 35–48. https://doi.org/10.1017/S0272263100011657

Lee, H. (2016). Concept-based instruction: Imagistic and metaphorical understanding of phrasal verbs. *English Teaching*, *71*(4), 167–191. https://doi.org/10.15858/engtea.71.4.201612.167

Lee, J.-A. (2001). Korean speakers. In M. Swan & B. Smith (Eds.), *Learner English: A teacher's guide to interference and other problems* (pp. 325–342). Cambrigde: Cambridge University Press.

Lee, J.-w., & Schallert, D. L. (1997). The relative contribution of L2 language proficiency and L1 reading ability to L2 reading performance: A test of the threshold hypothesis in an EFL context. *TESOL Quarterly*, *31*(4), 713–739. https://doi.org/10.2307/3587757

Lee, J.-Y. (2015). The use of English phrasal verbs in American spoken corpora. *International Journal of Language Studies*, *4*(9), 27–48.

Lee, S.-K. (2007). Effects of textual enhancement and topic familiarity on Korean EFL students' reading comprehension and learning of passive form. *Language Learning*, *57*(1), 87–118. https://doi.org/10.1111/j.1467-9922.2007.00400.x

Lee, S.-K., & Huang, H.-T. (2008). Visual input enhancement and grammar learning: A meta-analytic review. *Studies in Second Language Acquisition*, *30*, 307–331. https://doi.org/10.1017/S0272263108080479

Leeson, H. V. (2006). The mode effect: A literature review of human and technological issues in computerized testing. *International Journal of Testing*, *6*(1), 1-24. https://doi.org/10.1016/j.invent.2016.06.006

Lehmann, J., Goussios, C., & Seufert, T. (2016). Working memory capacity and disfluency effect: An aptitude-treatment-interaction study. *Metacognition and Learning*, *11*(1), 89–105. https://doi.org/10.1007/s11409-015-9149-z

Leidheiser, W., Branyon, J., Baldwin, N., Pak, R., & McLaughlin, A. (2015). Lessons learned in adapting a lab-based measure of working memory capacity for the web. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *59*(1), 756-760. https://doi.org/10.1177/1541931215591235

Lemons, C. J., King, S. A., Davidson, K. A., Puranik, C. S., Al Otaiba, S., & Fidler, D. J. (2017). Personalized reading intervention for children with Down syndrome. *Journal of School Psychology*. https://doi.org/10.1016/j.jsp.2017.07.006

Leow, R. P. (1997). The effects of input enhancement and text length on adult L2 readers' comprehension and intake in second language acquisition. *Applied Language Learning*, *8*(2), 151–82.

Leow, R. P. (2001). Do learners notice enhanced forms while interacting with the L2?: An online and offline study of the role of written input enhancement in L2 reading. *Hispania*, 496–509. https://doi.org/10.2307/3657810

Leow, R. P. (2009). Input enhancement and L2 grammatical development: What the research reveals. In S. L. Katz & J. Watzinger-Tharp (Eds.), *Conceptions of L2 grammar: Theoretical approaches and their applications in the L2 classroom.* Boston, MA: Heinle Publishers.

Leow, R. P. (2015). *Explicit learning in the L2 classroom: A student-centered approach.* New York: Routledge.

Leow, R. P., Egi, T., Nuevo, A. M., & Tsai, Y.-C. (2003). The roles of textual enhancement and type of linguistic item in adult L2 learners' comprehension and intake. *Applied Language Learning*, *13*(2), 1–16.

Leow, R. P., Hsieh, H.-C., & Moreno, N. (2008). Attention to form and meaning revisited. *Language Learning*, *58*(3), 665–695. https://doi.org/10.1111/j.1467-9922.2008.00453.x

Leow, R. P., & Martin, A. (2018). Enhancing the input to promote salience of the L2. In P. Spinner, S. M. Gass, & J. Behney (Eds.), *Salience in second language acquisition* (pp. 167–186). New York: Routledge.

Leow, R. P., & Zamora, C. C. (2017). Instructed second language acquisition (ISLA). In S. Loewen & M. Sato (Eds.), *The Routledge handbook of instructed second language acquisition* (pp. 33–49). New York: Routledge.

Levy, M. (2009). Technologies in use for second language learning. *The Modern Language Journal*, *93*(s1), 769–782. https://doi.org/10.1111/j.1540-4781.2009.00972.x

Li, S. (2015). The associations between language aptitude and second language grammar acquisition: A meta-analytic review of five decades of research. *Applied Linguistics*, *36*(3), 385–408. https://doi.org/10.1093/applin/amu054

Li, S. (2016). The construct validity of language aptitude: A meta-analysis. *Studies in Second Language Acquisition*, *38*(4), 801–842. https://doi.org/10.1017/S027226311500042X

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

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* (pp. 44–70). Amsterdam: John Benjamins.

Li, S., & Fu, M. (2016). Strategic and unpressured within-task planning and their associations with working memory. *Language Teaching Research.* https://doi.org/10.1177/1362168816684367

Li Shuay. (2017). An exploratory study on the role of foreign language aptitudes in instructed pragmatics learning in L2 Chinese. *Chinese as a Second Language Research*, *6*(1), 103–128. https://doi.org/10.1515/caslar-2017-0005

Liao, Y., & Fukuya, Y. J. (2004). Avoidance of phrasal verbs: The case of Chinese learners of English. *Language Learning*, *54*(2), 193–226. https://doi.org/10.1111/j.1467-9922.2004.00254.x

Lightbown, P. (1983). Exploring relationships between developmental and instructional sequences in L2 acquisition. In H. W. Seliger & M. H. Long (Eds.), *Classroom oriented research in second language acquisition* (pp. 217–243). Rowley, MA: Newbury House.

Linck, J. A. (2016). Analyzing individual differences in second language: The benefits of mixed effects models. In G. Granena, D. O. Jackson, & Y. Yilmaz (Eds.), *Cognitive individual differences in second language processing and acquisition* (pp. 105–128). John Benjamins Publishing.

Linck, J. A., & Cunnings, I. (2015). The utility and application of mixed-effects models in second language research. *Language Learning*, *65*(S1), 185–207. https://doi.org/10.1111/lang.12117

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

Linck, J. A., & Weiss, D. J. (2015). Can working memory and inhibitory control predict second language learning in the classroom? *SAGE Open*, *5*(4), 1–11. https://doi.org/10.1177/2158244015607352

Linnman, C., Carlbring, P., Åhman, Å., Andersson, H., & Andersson, G. (2006). The Stroop effect on the Internet. *Computers in Human Behavior*, *22*(3), 448–455.

https://doi.org/10.1016/j.chb.2004.09.010

Lipka, L. (1972). *Semantic structure and word-formation. Verb-particle constructions in contemporary English*. München: Fink.

Littlewood, W. (2011). Communicative language teaching: An expanding concept for a changing world. In E. Hinkel (Ed.), *Handbook of research in second language teaching and learning* (Vol. 2, pp. 541–557). London: Routledge.

Liu, C.-L., Wang, C.-H., Gao, Z.-M., & Huang, S.-M. (2005). Applications of lexical information for algorithmically composing multiple-choice cloze items. In J. Burstein & C. Leacock (Eds.), *Proceedings of the second workshop on building educational applications using NLP* (pp. 61–68). Ann Arbor, MI.

Liu, D. (2003). The most frequently used spoken American English idioms: A corpus analysis and its implications. *TESOL Quarterly*, *37*(4), 671–700. https://doi.org/10.2307/3588217

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

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. https://doi.org/10.1016/j.esp.2011.07.002

Loewen, S. (2011). Focus on form. In E. Hinkel (Ed.), *Handbook of research in second language teaching and learning* (Vol. 2, pp. 576–593). London: Routledge.

Loewen, S., & Inceoglu, S. (2016). The effectiveness of visual input enhancement on the noticing and L2 development of the Spanish past tense. *Studies in Second Language Learning and Teaching*, *1*(6), 89–110. https://doi.org/10.14746/ssllt.2016.6.1.5

Loewen, S., & Sato, M. (2017). Instructed second language acquisition (ISLA). In S. Loewen & M. Sato (Eds.), *The Routledge handbook of instructed second language acquisition* (pp. 1–12). New York: Routledge.

Long, M. H. (1983). Does second language instruction make a difference? A review of research. *TESOL Quarterly*, *17*(3), 359–382. https://doi.org/10.2307/3586253

Long, M. H. (1988). Instructed interlanguage development. In L. M. Beebe (Ed.), *Issues in second language acquisition: Multiple perspectives* (pp. 115–141). Rowley, MA: Newbury House.

Long, M. H. (1991). Focus on form: A design feature in language teaching methodology. In K. de Bot, R. B. Ginsberg, & C. Kramsch (Eds.), *Foreign language research in cross-cultural perspective* (pp. 39–52). Amsterdam: John Benjamins.

Long, M. H., & Robinson, P. (1998). Focus on form: Theory, research, and practice. In C. J. Doughty & J. Williams (Eds.), *Focus on form in classroom second language acquisition.* Cambridge: Cambridge University Press.

Loschky, L. (1994). Comprehensible input and second language acquisition: What is the relationship? *Studies in Second Language Acquisition*, *16*(3), 303–323. https://doi.org/10.1017/S0272263100013103

Loumbourdi, L. (2018). Communicative language teaching (CLT). In J. I. Liontas (Ed.), *The TESOL encyclopedia of English language teaching.* John Wiley & Sons, Inc. https://doi.org/10.1002/9781118784235.eelt0167

Lu, X. (2018). Natural language processing and intelligent computer-assisted language learning (ICALL). In J. I. Liontas (Ed.), *The TESOL encyclopedia of English language teaching.* Chichester, UK: Wiley Blackwell. https://doi.org/10.1002/9781118784235.eelt0422

Lum, J. A. G., Conti-Ramsden, G., Page, D., & Ullman, M. T. (2012). Working, declarative and procedural memory in specific language impairment. *Cortex*, *48*(9), 1138–1154. https://doi.org/10.1016/j.cortex.2011.06.001

Lyster, R. (2017). Content-based langauge teaching. In S. Loewen & M. Sato (Eds.), *The Routledge handbook of instructed second language acquisition* (pp. 87–107). New York: Routledge.

Mackey, A. (2017). Classroom-based research. In S. Loewen & M. Sato (Eds.), *The Routledge handbook of instructed second language acquisition* (pp. 541–561). New

York: Routledge.

Mackey, A., Philp, J., Egi, T., Fujii, A., & Tatsumi, T. (2002). Individual differences in working memory, noticing of interactional feedback and L2 development. In P. Robinson (Ed.), *Individual differences and instructed language learning* (pp. 181–210). Amsterdam: John Benjamins.

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

Magnusson, J. A., & Graham, C. R. (2011). The effect of explicit instruction versus exposure only on ESL learners' ability to recognize, use and recall phrasal verbs. *Applied Language Learning*, *21*(1-2), 29–52.

Mahpeykar, N., & Tyler, A. (2015). A principled cognitive linguistics account of English phrasal verbs with up and out. *Language and Cognition*, *7*(1), 1–35. https://doi.org/10.1017/langcog.2014.15

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

Manning, C. D. (2011). Part-of-speech tagging from 97% to 100%: Is it time for some linguistics? In *Computational Linguistics and Intelligent Text Processing. CICLing 2011. Lecture Notes in Computer Science, vol 6608* (pp. 171–189). Berlin/Heidelberg: Springer.

Martin, K. I., & N. C. Ellis. (2012). The roles of phonological short-term memory and working memory in L2 grammar and vocabulary learning. *Studies in Second Language Acquisition*, *34*(3), 379–413. https://doi.org/10.1017/S0272263112000125

Martínez-Flor, A., & Fukuya, Y. J. (2005). The effects of instruction on learners' production of appropriate and accurate suggestions. *System*, *33*(3), 463–480. https://doi.org/10.1016/j.system.2005.06.007

Matuschek, H., Kliegl, R., Vasishth, S., Baayen, H., & Bates, D. (2017). Balancing Type

I error and power in linear mixed models. *Journal of Memory and Language*, *94*, 305–315. https://doi.org/10.1016/j.jml.2017.01.001

McArthur, T. (1989). The long-neglected phrasal verb. *English Today*, *5*(2), 38–44. https://doi.org/10.1017/S026607840000393X

McCarthy, M., & O'Dell, F. (2004). *English phrasal verbs in use intermediate*. Cambridge: Cambridge University Press.

McDonald, A. S. (2002). The impact of individual differences on the equivalence of computer-based and paper-and-pencil educational assessments. *Computers & Education*, *39*(3), 299–312. https://doi.org/10.1016/S0360-1315(02)00032-5

McEnery, T., & Hardie, A. (2011). *Corpus linguistics: Method, theory and practice*. Cambridge: Cambridge University Press.

McLaughlin, B. (1980). Theory and research in second language learning: An emerging paradigm. *Language Learning*, *30*(2), 331–350. https://doi.org/10.1111/j.1467-1770.1980.tb00322.x

Meara, P. (2005). LLAMA language aptitude tests [Computer software manual]. Swansea, UK.

Medina, A., Callender, A. A., Brantmeier, C., & Schultz, L. (2017). Inserted adjuncts, working memory capacity, and L2 reading. *System*, *66*, 69–86. https://doi.org/10.1016/j.system.2017.03.002

Meguro, Y. (2017). Textual enhancement, grammar learning, reading comprehension, and tag questions. *Language Teaching Research*. https://doi.org/10.1177/1362168817714277

Melka, F. (1997). Receptive vs. productive aspects of vocabulary. In N. Schmitt & M. McCarthy (Eds.), *Vocabulary: Description, acquisition and pedagogy* (pp. 84–102). Cambridge: Cambridge University Press.

Meurers, D. (2013). Natural language processing and language learning. In *The encyclopedia of applied linguistics.* https://doi.org/10.1002/9781405198431.wbeal0858

Meurers, D. (2015). Learner corpora and natural language processing. In S. Granger, G. Gilquin, & F. Meunier (Eds.), *The Cambridge handbook of learner corpus research* (pp. 537–566). Cambridge: Cambridge University Press.

Meurers, D., & Dickinson, M. (2017). Evidence and interpretation in language learning research: Opportunities for collaboration with computational linguistics. *Language Learning*, *67*(S1), 66–95. https://doi.org/10.1111/lang.12233

Meurers, D., Ziai, R., Amaral, L., Boyd, A., Dimitrov, A., Metcalf, V., & Ott, N. (2010). Enhancing Authentic Web Pages for Language Learners. In *Proceedings of the 5th workshop on innovative use of NLP for Building Educational Applications (BEA) at NAACL-HLT 2010* (pp. 10–18). Los Angeles, California: Association for Computational Linguistics. Retrieved from http://aclweb.org/anthology/W10-1002.pdf

Mitchell, A. E., Jarvis, S., O'Malley, M., & Konstantinova, I. (2015). Working memory measures and L2 proficiency. In Z. Wen, M. Borges Mota, & A. McNeill (Eds.), *Working memory in second language acquisition and processing* (pp. 270–284). Bristol: Multilingual Multilingual Matters.

Miyake, A., & Friedman, N. P. (2012). The nature and organization of individual differences in executive functions: Four general conclusions. *Current Directions in Psychological Science*, *21*(1), 8–14. https://doi.org/10.1177/0963721411429458

Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., Howerter, A., & Wager, T. D. (2000). The unity and diversity of executive functions and their contributions to complex "frontal lobe" tasks: A latent variable analysis. *Cognitive Psychology*, *41*(1), 49–100. https://doi.org/10.1006/cogp.1999.0734

Mohammadi, E., & Mirdehghan, S. S. (2014). A CMC approach to teaching phrasal-verbs to Iranian EFL senior high school students: The case of blended learning. *Procedia*, *98*, 1174–1178. https://doi.org/10.1016/j.sbspro.2014.03.531

Mohammadi, E. N., & Tashakori, S. (2015). The effect of concordance-based approach on learning phrasal verbs to Iranian EFL learners. *International Journal of Research Studies in Language Learning*, *4*(3), 33–45. https://doi.org/10.5861/ijrsll.2014.910

Monfared, A., Cervantes, S. E., Lee, S. M., & Jackson, M. (2018). Establishing a praxis between SLA theory and CALL-based practices. In J. Perren, K. Kelch, J.-s. Byun, S. Cervantes, & S. Safavi (Eds.), *Applications of CALL theory in ESL and EFL environments* (pp. 59–79). Hershey, PA: IGI Global.

Moore, P. J. (2018). Task-Based Language Teaching (TBLT). In J. I. Liontas (Ed.), *The TESOL encyclopedia of English language teaching.* John Wiley & Sons, Inc. https://doi.org/10.1002/9781118784235.eelt0175

Morgan, P. S. (1997). Figuring out *figure out*: Metaphor and the semantics of English verb-particle construction. *Cognitive Linguistics*, *8*(4), 327–357. https://doi.org/10.1515/cogl.1997.8.4.327

Morgan-Short, K., Deng, Z., Brill-Schuetz, K. A., Faretta-Stutenberg, M., Wong, P. C., & Wong, F. C. (2015). A view of the neural representation of second language syntax through artificial language learning under implicit contexts of exposure. *Studies in Second Language Acquisition*, *37*(2), 383–419. https://doi.org/10.1017/S0272263115000030

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

Morgan-Short, K., Heil, J., Botero-Moriarty, A., & Ebert, S. (2012). Allocation of attention to second language form and meaning: Issues of think-alouds and depth of processing. *Studies in Second Language Acquisition*, *34*(4), 659–685. https://doi.org/10.1017/S027226311200037X

Morgan-Short, K., & Ullman, M. T. (2012). The neurocognition of second language. In S. M. Gass & A. Mackey (Eds.), *The Routledge handbook of second language acquisition* (pp. 282–299). New York: Routledge.

Murakami, A. (2013). Cross-linguistic influence on the accuracy order of L2 English grammatical morphemes. In S. Granger, G. Gilquin, & F. Meunier (Eds.), *Twenty years of learner corpus research. looking back, moving ahead: Proceedings 1* (pp. 325–334).

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

N. C. Ellis, & Sagarra, N. (2010). Learned attention effects in L2 temporal reference: The first hour and the next eight semesters. *Language Learning*, *60*, 85–108. https://doi.org/10.1111/j.1467-9922.2010.00602.x

N. C. Ellis, & Schmidt, R. (1998). Rules or associations in the acquisition of morphology? The frequency by regularity interaction in human and PDP learning of morphosyntax. *Language and Cognitive Processes*, *13*(2-3), 307–336. https://doi.org/10.1080/016909698386546

Nagata, N. (1996). Computer vs. workbook instruction in second language acquisition. *CALICO Journal*, 53–75.

Nagata, N. (2009). Robo-Sensei's NLP-based error detection and feedback generation. *CALICO Journal*, *26*(3), 562–579.

Nassaji, H. (2016). Form-focused instruction and second language acquisition. *Language Teaching*, *49*(1), 35–62. https://doi.org/10.1017/S0261444815000403

Nassaji, H. (2017). Grammar acquisition. In S. Loewen & M. Sato (Eds.), *The Routledge handbook of instructed second language acquisition* (pp. 205–223). New York: Routledge.

Nassaji, H., & Tian, J. (2010). Collaborative and individual output tasks and their effects on learning English phrasal verbs. *Language Teaching Research*, *14*(4), 397–419. https://doi.org/10.1177/1362168810375364

Nation, I. S. P. (2001). *Learning vocabulary in another language*. Cambridge: Cambridge University Press.

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

Nerbonne, J. (2003). Natural language processing in computer-assisted language learning. In R. Mitkov (Ed.), *The Oxford handbook of computational linguistics* (pp. 670–98). Oxford: Oxford University Press.

Nguyen, M. T. T., Pham, H. T., & Pham, T. M. (2017). The effects of input enhancement and recasts on the development of second language pragmatic competence. *Innovation in Language Learning and Teaching*, *11*(1), 45–67. https://doi.org/10.1080/17501229.2015.1026907

Norris, J. M., & Ortega, L. (2000). Effectiveness of L2 instruction: A research synthesis and quantitative meta-analysis. *Language Learning*, *50*(3), 417–528. https://doi.org/10.1111/0023-8333.00136

Nyns, R. R. (1989). Is intelligent computer-assisted language learning possible? *System*, *17*(1), 35–47. https://doi.org/10.1016/0346-251X(89)90058-4

O'Brien, I., Segalowitz, N., Freed, B., & Collentine, J. (2007). Phonological memory predicts second language oral fluency gains in adults. *Studies in Second Language Acquisition*, *29*(4), 557–581. https://doi.org/10.1017/S027226310707043X

Overstreet, M. (1998). Text enhancement and content familiarity: The focus of learner attention. *Spanish Applied Linguistics*, *2*(2), 229–258.

Pae, T.-I. (2018). A simultaneous analysis of relations between L1 and L2 skills in reading and writing. *Reading Research Quarterly*. https://doi.org/10.1002/rrq.216

Pam, P., & Karimi, L. (2016). The effect of textual enhancement technique on incidental learning of idiomatic expressions of Iranian intermediate students. *Theory and Practice in Language Studies*, *6*(5), 1121–1127. https://doi.org/10.17507/tpls.0605.28

Papaefthymiou-Lytra, S. C. (2001). Greek speakers. In M. Swan & B. Smith (Eds.), *Learner English: A teacher's guide to interference and other problems* (pp. 129–

144). Cambrigde: Cambridge University Press.

Paradis, M. (2004). *A neurolinguistic theory of bilingualism.* Amsterdam: John Benjamins.

Paradis, M. (2009). *Declarative and procedural determinants of second languages*. Amsterdam: John Benjamins.

Park, E. S. (2013). Learner-generated noticing behavior by novice learners: Tracing the effects of learners' L1 on their emerging L2. *Applied Linguistics*, *1*(34), 74–98. https://doi.org/10.1093/applin/ams016

Park, E. S., & Nassif, L. (2014). Textual enhancement of two L2 Arabic forms: A classroom-based study. *Language Awareness*, *23*(4), 334–352. https://doi.org/10.1080/09658416.2013.808645

Park, H., Choi, S., & Lee, M. (2012). Visual input enhancement, attention, grammar learning, & reading comprehension: An eye movement study. *English Teaching*, *67*(4). https://doi.org/10.15858/engtea.67.4.201212.241

Pasban, M. A., Forghani, M., & Nouri, A. (2015). The effects of using English captions on Iranian intermediate EFL students learning of phrasal verbs. *Journal of Language Teaching and Research*, *6*(2), 373–382.

Paulmann, S., Ghareeb-Ali, Z., & Felser, C. (2015). Neurophysiological markers of phrasal verb processing: Pvidence from L1 and L2 speakers. In R. R. Heredia & A. B. Cieślicka (Eds.), *Bilingual figurative language processing.* Cambridge: Cambridge University Press.

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

Peng, P., Barnes, M., Wang, C., Wang, W., Li, S., Swanson, H. L., ... Tao, S. (2017). A meta-analysis on the relation between reading and working memory. *Psychological Bulletin*, *1*(144), 48–76. https://doi.org/10.1037/bul0000124

Perrachione, T. K., Lee, J., Ha, L. Y., & Wong, P. C. (2011). Learning a novel phonological contrast depends on interactions between individual differences and training paradigm design. *The Journal of the Acoustical Society of America*, *130*(1), 461–472. https://doi.org/10.1121/1.3593366

Peters, E. (2012). Learning German formulaic sequences: The effect of two attention-drawing techniques. *The Language Learning Journal*, *40*(1), 65–79. https://doi.org/10.1080/09571736.2012.658224

Peters, E., Hulstijn, J. H., Sercu, L., & Lutjeharms, M. (2009). Learning L2 German vocabulary through reading: The effect of three enhancement techniques compared. *Language Learning*, *59*(1), 113–151. https://doi.org/10.1111/j.1467-9922.2009.00502.x

Petersen, K., & Sachs, R. (2015). The language classroom in the age of networked learning. In R. Leow, L. Cerezo, & M. Baralt (Eds.), *Technology and L2 learning: A psycholinguistic approach* (pp. 3–22). Berlin: De Gruyter Mouton.

Petersen, K. A. (2010). *Implicit corrective feedback in computer-guided interaction: Does mode matter?* (Doctoral dissertation). Georgetown University, Washington, DC.

Pfenninger, S. E., & Singleton, D. (2017). *Beyond age effects in instructional L2 learning: Revisiting the age factor*. Bristol: Multilingual Matters.

Pino-Silva, J. (1992). Extensive reading: No pain, no gain? *English Teaching Forum*, *30*(2), 48–49.

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

Plonsky, L. (2013). Study quality in SLA: An assessment of designs, analyses, and reporting practices in quantitative L2 research. *Studies in Second Language Acquisition*, *35*(4), 655–687. https://doi.org/10.1017/S0272263113000399

Plonsky, L. (2017). Quantitative research methods. In S. Loewen & M. Sato (Eds.), *The Routledge handbook of instructed second language acquisition* (pp. 505–521). New York: Routledge.

Plonsky, L., & Oswald, F. L. (2017). Multiple regression as a flexible alternative to

ANOVA in L2 research. *Studies in Second Language Acquisition*, *39*(3), 579–592. https://doi.org/10.1017/S0272263116000231

Poldrack, R. A., & Packard, M. G. (2003). Competition among multiple memory systems: Converging evidence from animal and human brain studies. *Neuropsychologia*, *41*(3), 245–251. https://doi.org/10.1016/S0028-3932(02)00157-4

Polio, C. (2007). A history of input enhancement: Defining an evolving concept. In C. Gascoigne (Ed.), *Assessing the impact of input enhancement in second language education: Evolution in theory, research and practice* (pp. 1–8). Stillwater, OK: New Forums Press.

Quené, H., & van den Bergh, H. (2008). Examples of mixed-effects modeling with crossed random effects and with binomial data. *Journal of Memory and Language*, *59*(4), 413–425. https://doi.org/10.1016/j.jml.2008.02.002

Quirk, R., Greenbaum, S., Leech, G., & Svartvik, J. (1985). *A comprehensive grammar of the English language*. London: Longman.

Quixal, M., & Meurers, D. (2016). How can writing tasks be characterized in a way serving pedagogical goals and automatic analysis needs? *CALICO Journal*, *33*(1), 19. https://doi.org/10.1558/cj.v33i1.26543

R Development Core Team. (2016). R: A language and environment for statistical computing [Computer software manual]. Vienna, Austria. Retrieved from http://www.R-project.org

Rahman, Z. A. A. A., & Abid, R. Z. (2014). Rarity or non-existence of phrasal verbs in the written discourse of Omani student-teachers of English. *SAGE Open*, *4*(4), 1-10. https://doi.org/10.1177/2158244014556988

Randal, H. (2012). Cognitive linguistics and the second language classroom. *TESOL Journal*, *46*(1), 6–29. https://doi.org/10.1002/tesq.5

Rankin, T. (2017). (Working) memory and L2 acquisition and processing. *Second Language Research*, *33*(3), 389-399. https://doi.org/10.1177/0267658316645387

Ranta, L., & Lyster, R. (2018). Form-focused instruction. In P. Garrett & J. M. Cots (Eds.), *The Routledge handbook of language awareness* (pp. 40–56). New York: Routledge.

Read, J. (1997). Vocabulary and testing. In N. Schmitt & M. McCarthy (Eds.), *Vocabulary: Description, acquisition and pedagogy.* Cambridge: Cambridge University Press.

Read, J. (2000). *Assessing vocabulary*. Cambridge: Cambridge University Press.

Reber, P. J., Knowlton, B. J., & Squire, L. R. (1996). Dissociable properties of memory systems: Differences in the flexibility of declarative and nondeclarative knowledge. *Behavioral Neuroscience*, *110*(5), 861–871. https://doi.org/10.1037/0735-7044.110.5.861

Reips, U.-D. (2002). Internet-based psychological experimenting: Five dos and five don’ts. *Social Science Computer Review*, *20*(3), 241–249. https://doi.org/10.1177/089443930202000302

Reips, U.-D., & Birnbaum, M. H. (2011). Behavioral research and data collection via the Internet. In R. W. Proctor & K.-P. L. Vu (Eds.), *The handbook of human factors in Web design* (pp. 563–585). Mahwah, New Jersey: Erlbaum.

Reips, U.-D., Buchanan, T., Krantz, J., & McGraw, K. (2015). Methodological challenges in the use of the Internet for scientific research: Ten solutions and recommendations. *Studia Psychologica*, *15*(2), 139–148. https://doi.org/10.21697/sp.2015.14.2.09

Resketi, B. M., & Bagheri, B. (2014). EFL learners’ incidental acquisition of English phrasal verbs through enhanced extensive reading vs. unenhanced extensive reading. *International Journal of Foreign Language Teaching and Research*, *2*(8), 33–42. Retrieved from http://jfl.iaun.ac.ir/article_12459_a0b1c9b6d4fec6e5ab8383f06051b56f.pdf

Révész, A., Michel, M., & Lee, M. (2017). Investigating IELTS academic writing task 2: Relationships between cognitive writing processes, text quality, and working memory. *IELTS Research Reports Online*

Series, 44. Retrieved from https://search.informit.com.au/documentSummary;dn=840774005843694;res=IELHSS

Reynolds, R., Schaf, E., & Meurers, D. (2014). A VIEW of Russian: Visual Input Enhancement and adaptive feedback. In *Proceedings of the 3rd Workshop on NLP for Computer-Assisted Language Learning at the 5th Swedish Language Technology Conference, Uppsala University, Sweden.*

Rhodes, S., & Cowan, N. (in press). Attention in working memory: Attention is needed but it yearns to be free. *Annals of the New York Academy of Sciences*.

Robinson, P. (1995). Attention, memory, and the "noticing" hypothesis. *Language Learning*, *45*(2), 283–331. https://doi.org/10.1111/j.1467-1770.1995.tb00441.x

Robinson, P. (1997). Individual differences and the fundamental similarity of implicit and explicit adult second language learning. *Language Learning*, *47*(1), 45–99. https://doi.org/10.1111/0023-8333.21997002

Robinson, P. (2001a). Individual differences, cognitive abilities, aptitude complexes and learning conditions in second language acquisition. *Second Language Research*, *17*(4), 368–392. https://doi.org/10.1177/026765830101700405

Robinson, P. (2001b). Task complexity, task difficulty, and task production: Exploring interactions in a componential framework. *Applied Linguistics*, *22*(1), 27–57. https://doi.org/10.1093/applin/22.1.27

Robinson, P. (2002). Effects of individual differences in intelligence, aptitude and workingmemory on incidental second language learning: A replication and extension of Reber, Walkenfield and Hernstadt (1991). In P. Robinson (Ed.), *Individual differences and instructed language learning* (pp. 211–266). Amsterdam: John Benjamins.

Robinson, P. (2005). Cognitive abilities, chunk-strength, and frequency effects in implicit artificial grammar and incidental L2 learning: Replications of Reber, Walkenfeld, and Hernstadt (1991) and Knowlton and Squire (1996) and their relevance for SLA. *Studies in Second Language Acquisition*, *27*(2), 235–268. https://doi.org/

10.1017/S0272263105050126

Robinson, P. (2007). Aptitudes, abilities, contexts and practice. In R. M. DeKeyser (Ed.), *Practice in a second language: Perspectives from applied linguistics and cognitive psychology* (pp. 256–286). Cambridge University Press.

Robinson, P. (2011). Second language task complexity, the cognition hypothesis, language learning, and performance. In P. Robinson (Ed.), *Second language task complexity. Researching the cognition hypothesis of language learning and performance* (pp. 3–38). Amsterdam: John Benjamins.

Robinson, P., Mackey, A., Gass, S. M., & Schmidt, R. (2012). Attention and awareness in second language acquisition. In *The Routledge handbook of second language acquisition* (pp. 247–267). New York: Routledge.

Rodríguez-Puente, P. (2012). The development of non-compositional meanings in phrasal verbs: A corpus-based study. *English Studies*, *93*(1), 71–90. https://doi.org/10.1080/0013838X.2011.638453

Roehr, K. (2008). Linguistic and metalinguistic categories in second language learning. *Cognitive Linguistics*, *19*(1), 67–106. https://doi.org/10.1515/COG.2008.005

Roehr, K. (2012). Aptitude treatment interaction (ATI) research. In P. Robinson (Ed.), *The Routledge encyclopedia of second language acquisition* (pp. 31–35). London: Routledge.

Rong, Y. (2015). A study of English phrase verb in language learning. *The Open Cybernetics & Systemics Journal*, *9*, 2128–2134.

Rosca, A., & de Altamirano, Y. B. (2016). Tracking down phrasal verbs: The case of *up* and *down*. *Onomázein*, *34*(18), 296–321. https://doi.org/10.7764/onomazein.34.18

Rudzewitz, B., Ziai, R., De Kuthy, K., & Meurers, D. (2017). Developing a web-based workbook for English supporting the interaction of students and teachers. In *Proceedings of the Joint 6th Workshopon NLP for Computer Assisted Language Learning and 2nd Workshop on NLP for Research on Language Acquisition at NoDaLiDa*

2017 (pp. 36–46).

Russell, V. (2012). Learning complex grammar in the virtual classroom: A comparison of processing instruction, structured input, computerized visual input enhancement, and traditional instruction. *Foreign Language Annals*, *45*(1), 42–71. https://doi.org/10.1111/j.1944-9720.2012.01168.x

Russell, V. (2014). A closer look at the output hypothesis: The effect of pushed output on noticing and inductive learning of the Spanish future tense. *Foreign Language Annals*, *47*(1), 25–47. https://doi.org/10.1111/flan.12077

Russell, V. (2016). An examination of learners' noticing and processing of complex Spanish grammar in authentic input texts. *Electronic Journal of Foreign Language Teaching*, *13*(1), 5–29.

Ryan, K., Hamrick, P., Miller, R. T., & Was, C. A. (2018). Salience, cognitive effort, and word learning: Insights from pupillometry. In S. M. Gass, P. Spinner, & J. Behney (Eds.), *Salience in second language acquisition* (pp. 187–200). New York: Routledge.

Ryoo, M.-L. (2013). A corpus-based study of the use of phrasal verbs in Korean EFL students' writing. *The Journal of Asia TEFL*, *10*(2), 63–89.

Sackett, P. R., Lievens, F., Van Iddekinge, C. H., & Kuncel, N. R. (2017). Individual differences and their measurement: A review of 100 years of research. *Journal of Applied Psychology*, *102*(3), 254. https://doi.org/10.1037/apl0000151

Sagarra, N. (2017). Longitudinal effects of working memory on L2 grammar and reading abilities. *Second Language Research*, *33*(3), 341–363. https://doi.org/10.1177/0267658317690577

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

Santamaria, K., & Sunderman, G. (2015). Working memory in processing instruction:

The acquisition of L2 French clitics. In Z. Wen, M. Borges Mota, & A. McNeill (Eds.), *Working memory in second language acquisition and processing* (pp. 205–223). Bristol: Multilingual Matters.

Sanz, C., & Grey, S. (2015). Effects of conditions on L2 development: Moving beyond accuracy. In P. Rebuschat (Ed.), *Implicit and explicit learning of languages* (pp. 301–324). Amsterdam/Philadelphia: John Benjamins Publishing.

Sanz, C., Lin, H.-J., Lado, B., Stafford, C. A., & Bowden, H. W. (2016). One size fits all? Learning conditions and working memory capacity in *Ab initio* language development. *Applied linguistics*, *37*(5), 669–692. https://doi.org/10.1093/applin/amu058

Sarab, M. R. A., & Kardoust, A. (2014). Concordance-based data-driven learning activities and learning English phrasal verbs in EFL classrooms. *Issues in Language Teaching*, *3*(1), 112–89. Retrieved from http://ilt.atu.ac.ir/article_1370.html

Sawyer, M., & Ranta, L. (2001). Aptitude, individual differences, and instructional design. In P. Robinson (Ed.), *Cognition and second language instruction* (pp. 319–353). Cambridge: Cambridge University Press.

Schenck, A. (2017). Learning to improve grammar instruction through comprehensive analysis of past research. *International Review of Applied Linguistics in Language Teaching*, *55*(2), 165–195. https://doi.org/10.1515/iral-2015-0038

Schmidt, R. (1990). The role of consciousness in second language learning. *Applied Linguistics*, *11*, 206–226. https://doi.org/10.1093/applin/11.2.129

Schmidt, R. (2001). Attention. In P. Robinson (Ed.), *Cognition and second language instruction* (pp. 3–32). Cambridge: Cambridge University Press. Retrieved from http://nflrc.hawaii.edu/PDFs/SCHMIDT%20attention.pdf

Schmidt, R. (2010). Attention, awareness, and individual differences in language learning. In W. M. Chan et al. (Eds.), *Proceedings of CLaSIC 2010, Singapore, December 2-4* (pp. 721–737). Singapore: National

University of Singapore, Centre for Language Studies. Retrieved from http://nflrc.hawaii.edu/PDFs/SCHMIDT%20Attention,%20awareness,%20and%20individual%20differences.pdf

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

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

Schneider, W., Eschman, A., & Zuccolotto, A. (2002). *E-prime: User's guide*. Pittsburgh, PA: Psychology Software Incorporated.

Schulenberg, S. E., & Yutrzenka, B. A. (1999). The equivalence of computerized and paper-and-pencil psychological instruments: Implications for measures of negative affect. *Behavior Research Methods, Instruments, & Computers*, *31*(2), 315–321. https://doi.org/10.3758/BF03207726

Schulze, M., & Heift, T. (2012). Intelligent CALL. In M. Thomas, H. Reinders, & M. Warschauer (Eds.), *Contemporary computer-assisted language learning* (pp. 249–265). London: Bloomsbury Academic.

Schunack, S. (2017). *Processing of non-canonical word orders in an L2* (Doctoral dissertation). University of Potsdam, Potsdam, Germany.

Schwartz, B. D. (1993). On explicit and negative data effecting and affecting competence and linguistic behavior. *Studies in Second Language Acquisition*, *15*(2), 47–163. https://doi.org/10.1017/S0272263100011931

Semmelmann, K., & Weigelt, S. (2017). Online psychophysics: Reaction time effects in cognitive experiments. *Behavior Research Methods*, *49*(4), 1241–1260. https://doi.org/10.3758/s13428-016-0783-4

Serafini, E. J., & Sanz, C. (2016). Evidence for the decreasing impact of cognitive ability on second language development as proficiency increases. *Studies in Second Language Acquisition*, *38*(4), 607–646. https://doi.org/10.1017/

S0272263115000327

Sharwood Smith, M. (1981). Consciousness-raising and second language acquisition theory. *Applied Linguistics*, *7*(3), 239–256.

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. Retrieved from 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. https://doi.org/10.1017/S0272263100011943

Sharwood Smith, M. (2013). Possibilities and limitations of enhancing language input: A MOGUL perspective. In A. G. Benati, C. Laval, & M. Arche (Eds.), *The grammar dimension in instructed second language learning* (pp. 36–57). London: Bloomsbury.

Sharwood Smith, M. (2017). Working with working memory and language. *Second Language Research*, *33*(3), 291–297. https://doi.org/10.1177/0267658317719315

Sheen, Y. (2007). The effect of focused written corrective feedback and language aptitude on ESL learners' acquisition of articles. *TESOL Quarterly*, *41*(2), 255–283. https://doi.org/10.1002/j.1545-7249.2007.tb00059.x

Sheen, Y. (2010). Differential effects of oral and written corrective feedback in the ESL classroom. *Studies in Second Language Acquisition*, *32*(2), 203–234. https://doi.org/10.1017/S0272263109990507

Shin, J.-A., & Christianson, K. (2012). Structural priming and second language learning. *Language Learning*, *62*(3), 931–964. https://doi.org/10.1111/j.1467-9922.2011.00657.x

Shipstead, Z., Harrison, T. L., & Engle, R. W. (2016). Working memory capacity and fluid intelligence: Maintenance and disengagement. *Perspectives on Psychological Science*, *11*(6), 771–799. https://doi.org/10.1177/1745691616650647

Shook, D. J. (1994). FL/L2 reading, grammatical information, and the input-to-intake phenomenon. *Applied Language Learning*, *5*(2), 57–93.

Shrout, P. E., & Rodgers, J. L. (2018). Psychology, science, and knowledge construction: Broadening perspectives from the replication crisis. *Annual Review of Psychology*, *69*(1), 487–510. https://doi.org/10.1146/annurev-psych-122216-011845

Side, R. (1990). Phrasal verbs: Sorting them out. *ELT Journal*, *2*(44), 144–152. https://doi.org/10.1093/elt/44.2.144

Simard, D. (2009). Differential effects of textual enhancement formats on intake. *System*, *37*(1), 124–135. https://doi.org/10.1016/j.system.2008.06.005

Sinclair, J. (1997). Corpus evidence in language description. In A. Wichmann, S. Fligelstone, T. McEnery, & G. Knowles (Eds.), *Teaching and language corpora* (pp. 27–39). London/New York: Longman.

Siyanova, A., & Schmitt, N. (2007). Native and nonnative use of multi-word vs. one-word verbs. *International Review of Applied Linguistics in Language Teaching*, *2*(45), 119–139. https://doi.org/10.1515/IRAL.2007.005

Sjöholm, K. (1995). *The influence of crosslinguistic, semantic, and input factors on the acquisition of English phrasal verbs: A comparison between Finnish and Swedish learners at an intermediate and advanced level* (Doctoral dissertation, Abo Akademi University Press). Retrieved from https://eric.ed.gov/?id=ED390305

Skehan, P. (1991). Individual differences in second language learning. *Studies in Second Language Acquisition*, *13*(2), 275–298. https://doi.org/10.1017/S0272263100009979

Skehan, P. (2012). Language aptitude. In *The Routledge handbook of second language acquisition* (pp. 381–395). New York: Routledge.

Skehan, P. (2016). Foreign language aptitude, acquisitional sequences, and psycholinguistic processes. In G. Granena, D. O. Jackson, & Y. Yilmaz (Eds.), *Cognitive* *individual differences in second language processing and acquisition* (pp. 17–40). Amsterdam: John Benjamins Publishing Company.

Skeppstedt, M., Ahltorp, M., & Henriksson, A. (2013). Vocabulary expansion by semantic extraction of medical terms. In *Proceedings of the Symposium on Languages in Biology and Medicine (LBM).*

Snow, R. E. (1991). Aptitude-treatment interaction as a framework for research on individual differences in psychotherapy. *Journal of Consulting and Clinical Psychology*, *59*(2), 205–216. https://doi.org/10.1037/0022-006X.59.2.205

Snow, R. E. (1992). Aptitude theory: Yesterday, today, and tomorrow. *Educational Psychologist*, *27*(1), 5–32. https://doi.org/10.1207/s15326985ep2701_3

Sonbul, S., & Schmitt, N. (2013). Explicit and implicit lexical knowledge: Acquisition of collocations under different input conditions. *Language Learning*, *63*(1), 121–159. https://doi.org/10.1111/j.1467-9922.2012.00730.x

Spada, N. (1997). Form-focused instruction and second language acquisition: A review of classroom and laboratory research. *Language Teaching*, *30*(2), 73–87. https://doi.org/10.1017/S0261444800012799

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

Spada, N., & Lightbown, P. M. (1999). Instruction, first language influence, and developmental readiness in second language acquisition. *The Modern Language Journal*, *1*(83), 1–22. https://doi.org/10.1111/0026-7902.00002

Spada, N., & Tomita, Y. (2010). Interactions between type of instruction and type of language feature: A meta-analysis. *Language Learning*, *60*(2), 263–308. https://doi.org/10.1111/j.1467-9922.2010.00562.x

Squire, L. R. (2004). Memory systems of the brain: A brief history and current perspective. *Neurobiology of Learning and Memory*, *82*(3), 171–177. https://

doi.org/10.1016/j.nlm.2004.06.005

Sternberg, R. J., Grigorenko, E. L., Ferrari, M., & Clinkenbeard, P. (1999). A triarchic analysis of an aptitude-treatment interaction. *European Journal of Psychological Assessment*, *15*(1), 3. https://doi.org/10.1027//1015-5759.15.1.3

Stewart, N., Chandler, J., & Paolacci, G. (2017). Crowdsourcing samples in cognitive science. *Trends in Cognitive Sciences*, *21*(10), 736–748. https://doi.org/10.1016/j.tics.2017.06.007

Stolarova, M., Brielmann, A. A., Wolf, C., Rinker, T., Burke, T., & Baayen, H. (2016). Early vocabulary in relation to gender, bilingualism, type, and duration of childcare. *Advances in Cognitive Psychology*, *12*(3), 130–144. https://doi.org/10.5709/acp-0192-6

Strong, B. (2013). A cognitive semantic approach to L2 learning of phrasal verbs. *The Language Teacher*, *37*(5), 28–31.

Sun, Y.-C., & Wang, L.-Y. (2003). Concordancers in the EFL classroom: Cognitive approaches and collocation difficulty. *Computer Assisted Language Learning*, *16*(1), 83–94. https://doi.org/10.1076/call.16.1.83.15528

Sung, M.-C. (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*. https://doi.org/10.1515/cllt-2017-0002

Sung, M.-C., & Kim, H. (2016). Tracing developmental changes in L2 learners' structuring of phrasal verbs: A corpus study of native and non-native argumentative essays. *3L: Southeast Asian Journal of English Language Studies*, *22*(2), 151–166. Retrieved from http://ejournal.ukm.my/3l/article/viewFile/11787/4383

Suzuki, Y., & DeKeyser, R. M. (2017). Exploratory research on second language practice distribution: An aptitude-by-treatment interaction. *Applied Psycholinguistics*, *38*(1), 27–56. https://doi.org/10.1017/S0142716416000084

Swain, M. (1985). Communicative competence: Some roles of comprehensible input and comprehensible output in its development. In S. M. Gass & C. G. Madden (Eds.), *Input and second language acquisition* (pp. 235–253). Rowley, Mss.: Newbury House.

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

Szudarski, P., & Carter, R. (2016). The role of input flood and input enhancement in EFL learners' acquisition of collocations. *International Journal of Applied Linguistics*, *26*(2), 245–265. https://doi.org/10.1111/ijal.12092

Tagarelli, K. M., Borges Mota, M., & Rebuschat, P. (2015). Working memory, learning context, and the acquisition of L2 syntax. In Z. Wen, M. Borges Mota, & A. McNeill (Eds.), *Working memory in second language acquisition and processing: Theory, research and commentary* (pp. 224–247). Bristol: Multilingual Matters.

Tagarelli, K. M., Ruiz, S., Moreno Vega, J. L., & 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), 293–316. https://doi.org/10.1017/S0272263116000036

Takashima, H., & Ellis, R. (1999). Output enhancement and the acquisition of the past tense. In R. Ellis (Ed.), *Learning a second language through interaction* (pp. 173–188). Amsterdam: John Benjamins.

Taylor, A., Marcus, M., & Santorini, B. (2003). The penn treebank: An overview. In A. Abeillé (Ed.), *Treebanks: Building and using parsed corpora* (pp. 5–22). New York: Springer.

Thibeau, T. J. (1999). *English prepositions in phrasal verbs: A study in second language acquisition* (Doctoral dissertation, The University of Arizona., Arizona, Arizona, United States) Retrieved from http://hdl.handle.net/10150/284018

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

Tolentino, L. C., & Tokowicz, N. (2014). Cross-language similarity modulates effectiveness of second language grammar instruction. *Language Learning*, *64*(2), 279–309. https://doi.org/10.1111/lang.12048

Tomlin, R. S., & Villa, V. (1994). Attention in cognitive science and second languageacquisition. *Studies in Second Language Acquisition*, *16*(2), 183–203. https://doi.org/10.1017/S0272263100012870

Torres-Martínez, S. (2017). Applied Cognitive Construction Grammar: A usage-based approach to the teaching of phrasal verbs (and other constructions). *European Journal of Applied Linguistics*. https://doi.org/10.1515/eujal-2016-0012

Trahan, D. E., & Larrabee, G. J. (1988). *Continuous Visual Memory Test*. Odessa, FL: Psychological Assessment Resources.

Trebits, A. (2009). The most frequent phrasal verbs in English language EU documents: A corpus-based analysis and its implications. *System*, *37*, 470–481. https://doi.org/10.1016/j.system.2009.02.012

Trofimovich, P., Lightbown, P. M., & Halter, R. (2013). Are certain types of instruction better for certain learners? *System*, *41*(4), 914–922. https://doi.org/10.1016/j.system.2013.09.004

Truscott, J. (2017). Modularity, working memory, and second language acquisition: A research program. *Second Language Research*, *33*(3), 313–323. https://doi.org/10.1177/0267658317696127

Turner, M. L., & Engle, R. W. (1989). Is working memory capacity task dependent? *Journal of Memory and Language*, *28*, 127–154. https://doi.org/10.1016/0749-596X(89)90040-5

Uchida, T. (2012). *Use of multiword verbs by non-advanced EFL learners: Focusing on common verb + particle combinations* (Vol. NBER Working Paper No. 8). Tokyo, Japan: Global COE program, Tokyo University of Foreign Studies. Retrieved from http://cblle.tufs.ac.jp/assets/files/publications/working_papers_08/section/303-323.pdf

Ullman, M. T. (2001). The neural basis of lexicon and grammar in first and second language: The declarative/procedural model. *Bilingualism: Language and Cognition*, *4*(2), 105–122. https://doi.org/10.1017/S1366728901000220

Ullman, M. T. (2004). Contributions of memory circuits to language: The declarative/procedural model. *Cognition*, *92*(1-2), 231–270. https://doi.org/10.1016/j.cognition.2003.10.008

Ullman, M. T. (2005). A cognitive neuroscience perspective on second language acquisition: The declarative/procedural model. In C. Sanz (Ed.), *Mind and context in adult second language acquisition: Methods, theory, and practice* (pp. 141–178). Georgetown University Press.

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

Ullman, M. T. (2016). The declarative/procedural model: A neurobiological model of language learning, knowledge, and use. In G. Hickok & S. A. Small (Eds.), *Neurobiology of language* (pp. 953–968). Amsterdam: Elsevier.

Ullman, M. T., & Lovelett, J. T. (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. https://doi.org/10.1177/0267658316675195

Ullman, M. T., & Pullman, M. Y. (2015). A compensatory role for declarative memory in neurodevelopmental disorders. *Neuroscience & Biobehavioral Reviews*, *51*, 205–222. https://doi.org/10.1016/j.neubiorev.2015.01.008

Unsworth, N., & Engle, R. W. E. (2007). The nature of individual differences in working memory capacity: Active maintenance in primary memory and controlled search from secondary memory. *Psychological Review*, *114*(1), 104.

Unsworth, N., Heitz, R. P., Schrock, J. C., & Engle, R. W. (2005). An automated version of the operation span task. *Behavior Research Methods*, *37*(3), 498–505. https://doi.org/10.3758/BF03192720

Vandenberg, R. J., & Lance, C. E. (2000). A review and synthesis of the measurement invariance literature: Suggestions, practices, and recommendations for organizational research. *Organizational Research Methods*, *3*(1), 4–70. https://doi.org/10.1177/109442810031002

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

VanPatten, B. (2012). Input processing. In *The Routledge handbook of second language acquisition* (pp. 268–281). New York: Routledge.

VanPatten, B., & Benati, A. G. (2015). *Key terms in second language acquisition* (2nd ed.). London: Bloomsbury.

VanPatten, B., & Cadierno, T. (1993). Explicit instruction and input processing. *Studies in Second Language Acquisition*, *15*(2), 225–243. https://doi.org/10.1017/S0272263100011979

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

Verspoor, M., Lowie, W., & Van Dijk, M. (2008). Variability in second language development from a dynamic systems perspective. *The Modern Language Journal*, *92*(2), 214–231. https://doi.org/10.1111/j.1540-4781.2008.00715.x

Villavicencio, A., Idiart, M. A., Ramisch, C., Araújo, V., Yankama, B., & Berwick, R. (2012). Get out but don't fall down: Verb-particle constructions in child language. In *Proceedings of the EACL 2012 Workshop on Computational Models of Language Acquisition and Loss* (pp. 43–50). Association for Computational Linguistics.

Waibel, B. (2007). *Phrasal verbs in learner English: A corpus-based study of German and Italian students* (Doctoral dissertation, Albert-Ludwigs-Universität, Freiburg, Germany). Retrieved from https://freidok.uni-freiburg.de/data/

3592

Waller, T. A. (2018). Content-based langauge teaching. In J. I. Liontas (Ed.), *The TESOL encyclopedia of English language teaching.* John Wiley & Sons, Inc. https://doi.org/10.1002/9781118784235.eelt0168

Walter, C. (2004). Transfer of reading comprehension skills to L2 is linked to mental representations of text and to L2 working memory. *Applied Linguistics*, *25*(3), 315–339. https://doi.org/10.1093/applin/25.3.315

Watcyn-Jones, P. (2001). *English phrasal verbs*. Edinburgh: Pearson Education Limited.

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

Weissheimer, J., & Borges Mota, M. (2009). Individual differences in working memory capacity and the development of L2 speech production. *Issues in Applied Linguistics*, *2*(17), 93–112. Retrieved from https://escholarship.org/uc/item/7tz0f9xw

Wen, Z. (2012). Working memory and second language learning. *International Journal of Applied Linguistics*, *22*(1), 1–22. https://doi.org/10.1111/j.1473-4192.2011.00290.x

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

Wen, Z. (2016). *Working memory and second language learning: Towards an integrated approach*. Bristol: Multilingual matters.

Wen, Z. (in press). Phonological and executive working memory as language aptitude. In Z. Wen, P. Skehan, A. Biedroń, S. Li, & R. Sparks (Eds.), *Language aptitude: Multiple perspectives and emerging trends.* Routledge.

Wen, Z., Biedroń, A., & Skehan, P. (2017). Foreign language aptitude theory: Yesterday, today and tomorrow. *Language Teaching*, *50*(1), 1–31. https://doi.org/

10.1017/S0261444816000276

Wen, Z., Borges Mota, M., & McNeill, A. (2015). *Working memory in second language acquisition and processing*. Bristol: Multilingual Matters.

Wesche, M. (1981). Language aptitude measures in streaming, matching students with methods, and diagnosis of learning problems. In K. C. Diller (Ed.), *Individual differences and universals in language learning aptitude* (pp. 119–154). Rowley, MA: Newbury House.

White, B. J. (2012). A conceptual approach to the instruction of phrasal verbs. *The Modern Language Journal*, *96*(3), 419–438. https://doi.org/10.1111/j.1540-4781.2012.01365.x

White, J. (2015). The effect of input-based instruction type on the acquisition of Spanish accusative clitics. *Hispania*, *98*(2), 264–284. https://doi.org/10.1353/hpn.2015.0032

Wickens, T. D. (2002). *Elementary signal detection theory*. Oxford: Oxford University Press.

Wilhelm, O., Hildebrandt, A. H., & Oberauer, K. (2013). What is working memory capacity, and how can we measure it? *Frontiers in Psychology*, *4*, 433. https://doi.org/10.3389/fpsyg.2013.00433

Williams, J. (1999). Learner-generated attention to form. *Language Learning*, *49*(4), 583–625. https://doi.org/10.1111/0023-8333.00103

Williams, J. N. (2012). Working memory and SLA. In P. Robinson, A. Mackey, S. M. Gass, & R. Schmidt (Eds.), *The Routledge handbook of second language acquisition* (pp. 427–441). New York: Routledge.

Williams, J. N. (2015). Working memory in SLA research: Challenges and prospects. In Z. Wen, M. Borges Mota, & A. McNeill (Eds.), *Working memory in second language acquisition and processing* (pp. 301–307). Bristol: Multilingual Matters.

Wilske, S. (2015). *Form and meaning in dialogue-based computer-assisted language learning* (Doctoral dissertation). Universität des Saarlandes, Saarbrücken.

Wilson, L., & Wilson, M. (2001). Farsi speakers. In M. Swan & B. Smith (Eds.), *Learner English: A teacher's guide to interference and other problems* (pp. 179–194). Cambrigde: Cambridge University Press.

Winke, P. M. (2013). The effects of input enhancement on grammar learning and comprehension: A modified replication of Lee (2007) with eye-movement data. *Studies in Second Language Acquisition*, *35*(2), 323–352. https://doi.org/10.1017/S0272263112000903

Wolfe, C. R. (2017). Twenty years of Internet-based research at SCiP: A discussion of surviving concepts and new methodologies. *Behavior Research Methods*, *49*(5), 1615–1620. https://doi.org/10.3758/s13428-017-0858-x

Wong, W. (2003). Textual enhancement and simplified input effects on L2 comprehension and acquisition of non-meaningful grammatical form. *Applied Language Learning*, *13*(2), 109–132.

Wong, W. (2004). Processing instruction in French: The roles of explicit information and structured input. In B. VanPatten (Ed.), *Processing instruction: Theory, research, and commentary* (pp. 187–205). Mahwah, New Jersey: Erlbaum.

Woodcock, R. W., McGrew, K. S., Mather, N., & Schrank, F. A. (2001). Woodcock-Johnson III tests of cognitive abilities examiner's manual [Computer software manual]. Itasca, IL.

Wray, A. (2006). Formulaic language. In K. Brown (Ed.), *Encyclopedia of language & linguistics* (2nd ed.). Amsterdam: Elsevier.

Wray, A. (2013). Formulaic language. *Language Teaching*, *46*(3), 316–334. https://doi.org/10.1017/S0261444813000013

Wyner, L., & Cohen, A. D. (2015). Second language pragmatic ability: Individual differences according to environment. *Studies in Second Language Learning and Teaching*, *5*(4), 519–556. https://doi.org/10.14746/ssllt.2015.5.4.2

Wyss, R. (2003). Putting phrasal verbs into perspective. *TESOL Journal*, *12*(1), 37–38. https://doi.org/10.1002/j.1949-3533.2003.tb00118.x

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

Yang, Y., Zhang, L. J., & Cheng, L. (2018). Effects of recasts and form-focused instruction on the acquisition of novel vocabulary. *Applied Linguistics Review*. https://doi.org/10.1515/applirev-2017-0084

Yasuda, S. (2010). Learning phrasal verbs through conceptual metaphors: A case of Japanese EFL learners. *TESOL Quarterly*, *44*(2), 250–273. https://doi.org/10.5054/tq.2010.219945

Yilmaz, Y. (2013). Relative effects of explicit and implicit feedback: The role of working memory capacity and language analytic ability. *Applied Linguistics*, *34*(3), 344-368. https://doi.org/10.1093/applin/ams044

Yilmaz, Y., & Granena, G. (2016). The role of cognitive aptitudes for explicit language learning in the relative effects of explicit and implicit feedback. *Bilingualism: Language and Cognition*, *19*(1), 147–161. https://doi.org/10.1017/S136672891400090X

Yoshitomi, A. (2006). The use of phrasal verbs by Japanese learners of English: Implications from story telling data. In A. Yoshitomi, T. Umino, & M. Negishi (Eds.), *Readings in second language pedagogy and second language acquisition: In Japanese context* (pp. 201–225). Amsterdam: John Benjamins Publishing.

Zabihi, R. (2018). The role of cognitive and affective factors in measures of L2 writing. *Written Communication*, *35*(1), 32–57. https://doi.org/10.1177/0741088317735836

Zettler, I. (2011). Self-control and academic performance: Two field studies on university citizenship behavior and counterproductive academic behavior. *Learning and Individual Differences*, *21*(1), 119–123. https://doi.org/10.1016/j.lindif.2010.11.002

Zhou, H., Rossi, S., & Chen, B. (2017). Effects of working memory capacity and tasks in processing L2 complex sentence: Evidence from Chinese-English bilinguals. *Frontiers in Psychology*, *8*, 595. https://doi.org/10.3389/fpsyg.2017.00595

Zhou, H., Rossi, S., Li, J., Liu, H., Chen, R., & Chen, B. (2017). Effects of working memory capacity in processing wh-extractions: Eye-movement evidence from Chinese–English bilinguals. *Journal of Research in Reading*, *40*(4), 420–438. https://doi.org/10.1111/1467-9817.12079

Ziegler, N., Meurers, D., Rebuschat, P., Ruiz, S., Moreno Vega, J. L., Chinkina, M., ... Grey, S. (2017). Interdisciplinary research at the intersection of CALL, NLP, and SLA: Methodological implications from an input enhancement project. *Language Learning*, *67*, 209–231. https://doi.org/10.1111/lang.12227

Zilio, L., Wilkens, R., & Fairon, C. (2017). Enhancing grammatical structures in web-based texts. *CALL in a climate of change: Adapting to turbulent global conditions–short papers from EUROCALL 2017*, 345–350. https://doi.org/10.14705/rpnet.2017.eurocall2017.738

Zyzik, E. (2012). Input-based instructional approaches. In C. A. Chapelle (Ed.), . Blackwell Publishing Ltd. https://doi.org/10.1002/9781405198431.wbeal0542

Appendix

1 Test items

1.1 Targets

	Test item	Answer
1.	In big action movies, bad guys usually ____ famous buildings. (explode)	blow up
2.	In my country, inflation has been ____ in the last years. (increase)	building up
3.	As a fan, he ____ the entire collection of his favorite painter. (acquire)	bought up
4.	The survey was ____ to determine children's eating habits. (complete)	carry out
5.	She made a genuine effort to ____ her company. (sanitize)	clean up
6.	Experts ____ with many strategies to improve the economy in our region. (produce)	came up
7.	After falling asleep on the train, he ____ in the wrong city. (finish)	ended up
8.	Doctors couldn't ____ what happened to the patient. (understand)	figure out
9.	Her mother ____ smoking after 25 years. (abandon)	gave up
10.	The teacher encouraged her students to ____ to the next chapter and see what happened. (proceed)	go on
11.	Renting a more expensive apartment ____ my savings. (consume)	gobbled up
12.	He enjoyed ____ in a small village in the mountains. (mature)	growing up
13.	The company has ____ over half of its workforce. (disemploy)	laid off
14.	The city has ____ plans for the construction of a new stadium. (present)	laid out
15.	Young people ____ the majority of the population in my country. (constitute)	make up
16.	The new shopping center ____ many opportunities for business. (expand)	opens up
17.	Lower oil prices make it difficult for some countries to ____ their debts. (liquidate)	pay off
18.	Some researchers have ____ that eating a balanced breakfast reduces obesity. (indicate)	pointed out
19.	The government has promised to ____ efforts to improve education. (intensify)	ramp up

	Test item	Answer
20.	She ____ the possibility of a complete reconciliation with her father. (exclude)	ruled out
21.	Firefighters intend to ____ the response to fires in the affected areas. (maximize)	scale up
22.	Candidates have ____ their strategy to reduce poverty . (present)	set out
23.	The university ____ a committee to investigate the incident. (establish)	set up
24.	During the 2008 economic crisis, many governments intervened to ____ their banks. (support)	shore up
25.	The police ____ several illegal nightclubs. (close)	shut down
26.	The government needs to ____ more volunteers to assist in rescue operations. (register)	sign up
27.	Only few people were willing to ____ the demanding job of an office manager. (accept)	take on
28.	She ____ the family business after her father retired. (assume)	took over
29.	When the strategy ____ to be effective, they adopted it. (prove)	turned out
30.	The court ordered a blogger to ____ her negative review of a local restaurant. (attenuate)	water down

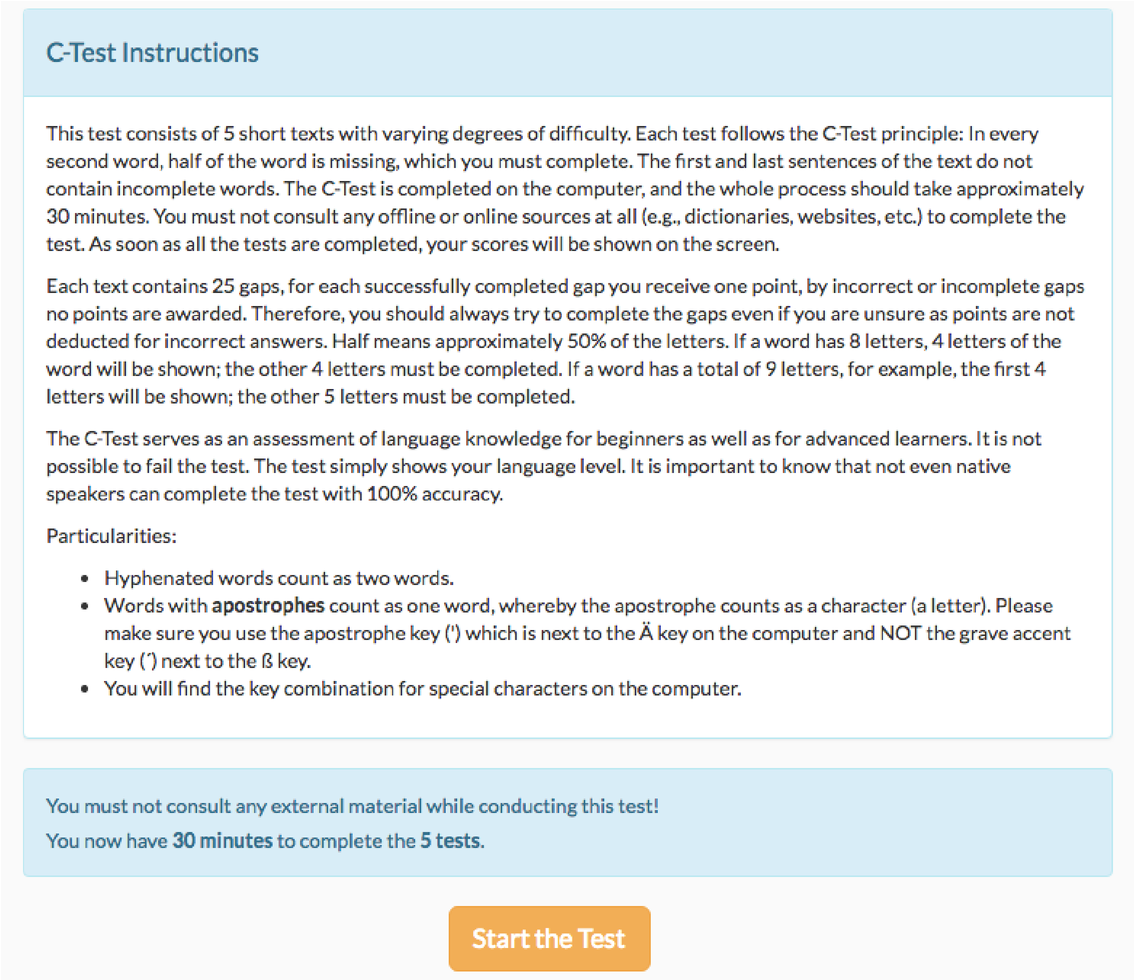
1.2 Distractors

	Test item	Answer
1.	The animal ____ from its cell and caused panic among people. (escape)	brooke loose
2.	Recent evidence ____ on the innocence of the accused. (discredit)	casts doubt
3.	Because he lost his job, the family had to ____ to pay the bills. (economize)	cut corners
4.	The mother saved her child from ____ in front of his classmates. (be humiliated)	losing face
5.	The team promised to ____ for last season's disappointments. (compensate)	make amends
6.	The comedian's offensive jokes ____ across the Internet. (cause trouble)	made waves
7.	The senior manager ____ and received special treatment. (use authority)	pulled rank
8.	She in order to obtain tickets for the exclusive fashion event. (use influence)	pulled strings
9.	Archaeologists have ____ to preserve the evidence of a new dinosaur species. (work carefully)	taken pains
10.	The new discovery ____ on the origins of human species. (explain)	throws light

2 Target corpus

Text number	(shortened) URL	Length[1]
1	https://reut.rs/2NTjEMv	737
2	https://reut.rs/2mkwxm5	1,686
3	https://reut.rs/2mfRoa2	3,291
4	https://reut.rs/2Nf3DPI	995
5	https://reut.rs/2LmPe3z	2,485
6	https://reut.rs/2mhUjPB	913
7	https://reut.rs/2uInq2v	938
8	https://reut.rs/2mkAozF	3,620
9	https://reut.rs/2L81mbJ	1,635
10	https://reut.rs/2NQpspP	882
11	https://tmsnrt.rs/2uDYhpE	2,631
12	https://reut.rs/2Lh3ESI	1,115
13	https://reut.rs/2mkqq0W	970
14	https://tmsnrt.rs/2NP7Ydn	1,427
15	https://reut.rs/2LbKFw6	1,321
16	https://tmsnrt.rs/2KVef9y	1,131
17	https://reut.rs/2uByYEq	1,293
18	https://reut.rs/2umxwqc	1,337
19	https://reut.rs/2NjIoMr	904
20	https://reut.rs/2mgANmC	907
21	https://reut.rs/2zCnXbP	1,471
22	https://reut.rs/2zE3fYZ	491
23	https://reut.rs/2Jp1XAJ	2,142
24	https://tmsnrt.rs/2NSWsxB	815
25	https://reut.rs/2Lj03Ud	758
26	https://reut.rs/2Lf9xTx	783
27	https://reut.rs/2JpeJiT	782
28	https://reut.rs/2uvmJKD	724
29	https://reut.rs/2LaBDiR	1,012
30	https://reut.rs/2uCLzao	1,809
31	https://reut.rs/2NUvhCu	1,140
32	https://reut.rs/2L11I4w	909
33	https://reut.rs/2NSs32H	711
34	https://reut.rs/2urFUVm	734
35	https://reut.rs/2uqMVFU	945
36	https://reut.rs/2mgDowS	656
37	https://tmsnrt.rs/2Jqo1uG	617
38	https://reut.rs/2NfMTaY	741
39	https://reut.rs/2uwaAUA	565
40	https://reut.rs/2uojeWd	984

Note. [1] = tokens.


C-Test Instructions

This test consists of 5 short texts with varying degrees of difficulty. Each test follows the C-Test principle: In every second word, half of the word is missing, which you must complete. The first and last sentences of the text do not contain incomplete words. The C-Test is completed on the computer, and the whole process should take approximately 30 minutes. You must not consult any offline or online sources at all (e.g., dictionaries, websites, etc.) to complete the test. As soon as all the tests are completed, your scores will be shown on the screen.

Each text contains 25 gaps, for each successfully completed gap you receive one point, by incorrect or incomplete gaps no points are awarded. Therefore, you should always try to complete the gaps even if you are unsure as points are not deducted for incorrect answers. Half means approximately 50% of the letters. If a word has 8 letters, 4 letters of the word will be shown; the other 4 letters must be completed. If a word has a total of 9 letters, for example, the first 4 letters will be shown; the other 5 letters must be completed.

The C-Test serves as an assessment of language knowledge for beginners as well as for advanced learners. It is not possible to fail the test. The test simply shows your language level. It is important to know that not even native speakers can complete the test with 100% accuracy.

Particularities:

- Hyphenated words count as two words.
- Words with **apostrophes** count as one word, whereby the apostrophe counts as a character (a letter). Please make sure you use the apostrophe key (') which is next to the Ä key on the computer and NOT the grave accent key (`) next to the ß key.
- You will find the key combination for special characters on the computer.

You must not consult any external material while conducting this test!

You now have **30 minutes** to complete the **5 tests**.

Start the Test

3 Instructions

3.1 C-test

3.2 Pretest and posttest language tests

At the beginning of the test

It is important that you observe the following instructions:

- Take the test in a quiet place without interruption.
- Complete the test from start to finish in one sitting.
- The program will not allow to go back or retake any of the tasks, so do not use the browser's back button.
- Do not refresh or close the browser window.

Very important: During the test, please do not take notes, consult with others, books or the Internet.

Immediately after the test

Answer the following questions:

1. Did you follow the instructions exactly as described at the beginning of the test?
2. Would you recommend including your data in our analysis?

3.3 Block of cognitive tests

At the beginning of the block

It is important that you observe the following instructions:

- Do the experiment in a quiet place without interruption.
- Complete the experiment from start to finish in one sitting.
- The program will not allow to go back or retake any of the tasks, so do not use the browser's back button.
- Do not refresh the browser page.
- Do not close the browser window.

Very important: During the test, please do not take notes, consult with others, books or the Internet.

3.4 Treatment

3.4.1 Meaning-focused condition

Please read at least 2 articles per day.

Please click on the title of an article from the REUTERS News Service to read. Once you have finished reading the article, please click on the "Click to continue to questionnaire" button at the bottom.

Remember: Please read at least 2 (up to 5) articles per day.

3.4.2 Form-focused condition

Please read at least 2 articles per day.

Please click on the title of an article from the REUTERS News Service to read.

Please read the article carefully. When reading the article, you will notice that there are several blank spaces (with dropdown menus) where normally words like “up”, “out” or “in” would appear next to the verb. Please click on the boxes and select the correct option from the pulldown menu.

Please select what you believe to be the correct answer! If you are stuck and really do not know the answer at all, then click on the smiley face next to the box. It will give you the correct answer. But please ONLY USE THIS, if you really do not know the answer.

Remember: Please read at least 2 (up to 5) articles per day.