LANGUAGE APTITUDE AND WORKING MEMORY IN THIRD LANGUAGE LEARNING

ELİFCAN ÖZTEKİN

BOĞAZİÇİ UNIVERSITY

2023

LANGUAGE APTITUDE AND WORKING MEMORY IN THIRD LANGUAGE LEARNING

Dissertation submitted to the

Institute for Graduate Studies in Social Sciences

in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

in

English Language Education

by

Elifcan Öztekin

Boğaziçi University

2023

DECLARATION OF ORIGINALITY

, Elifcan Öztekin, certify that

- I am the sole author of this thesis and that I have fully acknowledged and documented in my thesis all sources of ideas and words, including digital resources, which have been produced or published by another person or institution;
- this thesis contains no material that has been submitted or accepted for a degree or diploma in any other educational institution;
- this is a true copy of the thesis approved by my advisor and thesis committee at Boğaziçi University, including final revisions required by them.

Signature.....

Date .....

ABSTRACT

Language Aptitude and Working Memory in Third Language Learning

This study investigated the relationships of visuospatial and verbal working memory (WM) tasks with language aptitude (LA) and their predictive roles in explaining third language (L3) comprehension. The participants were 110 L1 Turkish-speaking majors in English language teaching, who studied various additional foreign languages for varying numbers of semesters at beginner, intermediate, and upper-intermediate / advanced levels, including a control group with no L3. They took the LLAMA version 3 (Meara & Rogers, 2019), forward digit span, rotation, symmetry, and operation span tasks (Foster et al., 2015). Participants learning one of the four designated European languages (i.e., German, French, Italian, and Spanish) also took reading and listening comprehension sections of general proficiency tests created by the certified cultural institutes. Principal component analysis (PCA) revealed a three-factor solution, interpreted as verbal/phonological memory (digit span, operation span, and partially LLAMA D), visuospatial WM (symmetry and rotation span), and LA (LLAMA B, E, F, partially D). Univariate analyses of variance on the three factors did not reveal any significant relationships with L3 experience. The hierarchical regression on L3 listening comprehension indicated that L3 experience and the LA factor cumulatively explained 45% of the variance, with the LA factor contributing an additional 15% of the variance. The same set of predictors did not explain any significant variance in L3 reading comprehension. The study confirmed the dissociation between certain LA and WM abilities, but some overlap in verbal/phonological/memory abilities. The distinctive LA factor was considerably influential for L3 listening but not reading comprehension.

ÖZET

Üçüncü Dil Öğreniminde Dile Yatkınlık ve İşler Bellek

Bu çalışma, görsel-uzaysal ve sözel işler bellek (İB) görevlerinin dile yatkınlık (DY) ile ilişkilerini ve bunların üçüncü dilde (D3) okuduğunu ve dinlediğini anlamayı açıklamadaki belirleyici rollerini araştırmıştır. Katılımcılar, başlangıç, orta ve üst-orta/ileri düzeylerde çeşitli ek yabancı diller öğrenen ve herhangi bir D3 öğrenmeyen bir kontrol grubu da dahil olmak üzere, İngilizce öğretmenliği anadalında öğrenim gören, anadili Türkçe olan 110 kişiden oluşmaktadır. Katılımcılar, LLAMA testleri sürüm 3 (Meara & Rogers, 2019), ileri sayı dizisi, döndürme, simetri ve işlem aralığı görevlerini (Foster ve diğerleri, 2015) tamamlamışlardır. Ayrıca, belirlenmiş dört Avrupa dilinden (Almanca, Fransızca, İtalyanca ve İspanyolca) birini öğrenen katılımcılar sertifikalı kültür enstitüleri tarafından oluşturulan genel yeterlilik sınavlarının okuma ve dinlediğini anlama bölümlerine de almışlardır. Temel bileşen analizi, sözel/fonolojik bellek (sayı aralığı, işlem aralığı ve LLAMA D), görsel-uzaysal İB (simetri ve döndürme aralığı) ve DY (LLAMA B, E, F, kısmen D) olarak yorumlanan 3 faktörlü bir çözüm ortaya çıkarmıştır. Bu üç faktör ile D3 deneyimi arasındaki ilişkiyi inceleyen tek değişkenli varyans analizleri, anlamlı olmayan sonuçlar ortaya çıkarmıştır. D3 dinlediğini anlamadaki hiyerarşik regresyon, D3 deneyimi ve DY faktörünün kümülatif olarak varyansın %45'ini açıkladığını ve DY faktörünün bu varyansın %15'ini açıkladığını göstermiştir. Aynı yordayıcı faktörler, D3 okuduğunu anlamada herhangi bir önemli varyans açıklamamıştır. Çalışma, belirli DY ve İB yetenekleri arasındaki ayrışmayı doğrulamış, ancak sözel/fonolojik hafıza yeteneklerinde bazı örtüşmeler göstermiştir. Ayrışmış olan DY faktörü, D3 dinlemede önemli ölçüde etkiliyken, okuduğunu anlamada etkisi gözlenmemiştir.

CURRICULUM VITAE

NAME: Elifcan Öztekin

DEGREES AWARDED
PhD in English Language Education, 2023, Boğaziçi University
MA in English Language Education, 2016, Boğaziçi University
BA in English Language Teaching, 2013, Boğaziçi University

AREAS OF SPECIAL INTEREST

Psycholinguistics, cognitive individual differences, language skills development, multilingualism, metacognition and strategy awareness, digital learning in language and culture

PROFESSIONAL EXPERIENCE
Research Assistant, Department of Foreign Language Education, Boğaziçi University, 2013 - present

AWARDS AND HONORS
Highest Honors List, Boğaziçi University, 2016 (GPA: 3.92)
Highest Honors List, Boğaziçi University, 2013 (GPA: 3.79)

GRANTS
TÜBİTAK PhD Scholarship, 2016 – 2020
TÜBİTAK MA Scholarship, 2013 – 2015

PUBLICATIONS
Journal Articles

Öztekin, E. & Candan, E. (Online first, November 2022). The role of language aptitude probed within extensive instruction experience: morphosyntactic knowledge of advanced users of L2 English. International Review of Applied Linguistics in Language Teaching (SSCI).

Öztekin, E. & Erçetin, G. (2022). Metacognitive awareness of listening strategies in second and third language: The role of language proficiency across and within languages. International Journal of Listening (Scopus), 36(3), 235-253.

Guerra, L., Cavalheiro, L., Pereira, R., Kurt, Y., Öztekin, E., Candan, E., & Bayyurt, Y. (2022). Representations of the English as a Lingua Franca framework: Identifying ELF-aware activities in Portuguese and Turkish coursebooks. RELC Journal (SSCI), 53(1), 134-150.

Öztekin, E., Özer, F., Belin, M., Büyüksolak, A., Ertaş, G., Mutlu Gülbak, G., Pesen, M., Şeker, V., & Türkmen, C. E. (2020). Öğretim üyeleri ve araştırma görevlileri perspektifinden fakültenin araştırma potansiyeli: Boğaziçi Üniversitesi Eğitim Fakültesi durumu. (Research potential of faculty from the perspectives of faculty members and research assistants: The case of Bogaziçi University Faculty of Education). Boğaziçi Üniversitesi Eğitim Dergisi (ULAKBİM), 37(2), 107-122.

Bayyurt, Y. Kurt, Y., Öztekin, E., Guerra, L., Cavalheiro, L. & Pereira, R. (2019). English language teachers' awareness of English as a Lingua Franca in multilingual and multicultural contexts. Eurasian Journal of Applied Linguistics (ESCI), 5(2), 185-202.

Öztekin, E., & Haznedar, B. (2015). Dillerarası etkileşim: Çokdilli ortamlarda sözdizimsel gelişim ve nesne düşürme. (Crosslinguistic influence: Syntactic development and object drop in multilingual contexts). Boğaziçi Üniversitesi Eğitim Dergisi, 32(2), 37-47. https://dergipark.org.tr/en/download/article-file/311765

Book Chapters

Öztekin, E. & Candan, E. (2019). The role of corpora in language teaching. In *Research trends in English language teacher education and English language teaching,* University of Évora, 327-342.

Conference Proceedings

Candan, E., Öztekin, E., Kurt, Y., Bayyurt, Y., Guerra, L., Cavalheiro, L., & Pereira, R. (2019). Representations of ELF in language teachers' beliefs and ELT coursebooks: Findings from Turkey and Portugal. In the 1st International ILTERG Conference Proceedings, Antalya, Turkey.

Öztekin, E. & Haznedar, B. (2015). Dillerarası etkileşim: Çokdilli ortamlarda sözdizimsel gelişim ve nesne düşürme. (Crosslinguistic influence: Syntactic development and object drop in multilingual contexts). In 29. National Linguistics Conference (Ulusal Dilbilim Kurultayı) Proceedings, Kocaeli University, Kocaeli, Turkey.

ACKNOWLEDGEMENTS

My deepest and greatest thanks go to my PhD supervisor, Prof. Gülcan Erçetin, who has always inspired me with her academic standing and personal integrity in all of her work. She has supported and guided me all through my academic journey, enlightening my way with her wisdom.

I would like to express my heartfelt gratitude to my dissertation supervisory committee members; Prof. Ayşe Gürel and Assist. Prof. Zeynep Çamlıbel Acar, who provided invaluable feedback on my work in progress for this dissertation. I had the chance to improve aspects of this dissertation project thanks to their insights. I would like to express my gratitude in particular to Prof. Gürel for the support and advice she provided whenever I needed it without any hesitance. I also would like to thank the dissertation committee members, Assist. Prof. Şebnem Yalçın and Assist. Prof. Mehmet Akıncı for their invaluable and constructive feedback on the final version of this dissertation. I also would like to thank all my friends, colleagues and department faculty for their support and understanding throughout my journey.

I also would like to express my sincere appreciation to the participants, who volunteered to spare their time and energy to contribute to this study. I must thank my dear family; my parents, Sevim and Satı Öztekin and my sister, Pınar Öztekin. They have supported and encouraged me throughout this journey. I found the strength to continue in their close company whenever I needed it.

Finally, I express my thanks to the Scientific and Technological Research Council of Turkey for the support I received for my doctoral study within the 2211-A PhD scholarship program.

TABLE OF CONTENTS

LIST OF TABLES

CHAPTER 1

INTRODUCTION

1.1 Background of the study

Research on individual differences in second language (L2) learning addressed various cognitive factors such as learner strategies, cognitive/learning style, working memory (WM), attention, and language aptitude (LA) (e.g. Dörnyei & Skehan, 2003; Granena, Jackson, & Yılmaz, 2016). Among these factors, LA and WM received attention regarding their nature and roles in explaining language attainment, while there has not yet been an established pattern on their construct-level structures supported with consistent evidence (Wen, Biedroń, & Skehan, 2017).

WM, basically defined as a limited capacity of storage and processing, was reported as a factor related to various aspects of L2 performance such as comprehension (e.g. Service, Simola, Metsanheimo, & Maury, 2002), learning grammar (e.g., Serafini & Sanz, 2016; Williams & Lovatt, 2003) and vocabulary (e.g., Martin & Ellis, 2013). Foreign language aptitude, on the other hand, was initially defined as a set of abilities that facilitated L2 learning in primarily instructed (but also naturalistic) environments with higher attainment rates at early stages (Carroll, 1962; 1990). Providing evidence of WM measured in the L2 explaining L2 syntactic comprehension and cue preference, Miyake and Friedman (1998) claimed that it could be WM functions underlying the abilities attributed to "language aptitude" back then, and they proposed "WM as language aptitude" idea. However, more recent accounts of LA emphasize a more multi-layered view of it rather than a unitary and monolithic structure (Robinson, 2005; Wen & Skehan, 2011) with some

aspects of WM and LA interacting and some being distinct. In this line, evidence from more recent research lends support for such a multi-layered view of aptitude in relation to WM, pointing to several interrelations as well as distinctions among components of LA and WM (e.g. Granena, 2013; Hummel, 2009; Sáfár & Kormos, 2008; Yalçın, Çeçen, & Erçetin, 2016). These findings are interpreted within the theoretical arguments such as the age-related effects in L2 learning regarding access to language-specific cognitive mechanisms. While some positions highlight that L2 learning processes rely more essentially on domain-general rather than language-specific mechanisms after a certain critical age (e.g., DeKeyser, 1997, 2000, 2019), some other positions propose that L2 learning still entails domain-specific language-related properties at any stage with changing weights at different stages of building L2 knowledge even after a critical period for L2 learning (e.g., Skehan, 2016, 2019). In this regard, LA could be argued to represent abilities specialized in language-specific abilities such as language analysis and phonemic coding, while WM functions represent domain-general abilities capable of processing visual as well as verbal input through management and storage capacities, effective in a wide range of cognitive functions. At this point, understanding the roles of these domain-general and domain-specific cognitive resources that LA and WM represent and their interactions is still an important research question in the field.

The role of these cognitive differences in relation to learning a third language (L3) received growing attention. Some L3 studies questioned whether the cognitive advantages of bilingualism observed as superior performance of bilinguals on tasks measuring executive functions such as inhibitory control, task switching or updating (e.g. Bialystok, 2009; 2011) compared to their monolingual counterparts can be extended to learning L3/n languages. This executive function aspect of WM in

managing various cognitive processes was particularly associated with an inherent bilingualism effect due to the competing representations of two languages, requiring a managerial system to effectively control the activation of the two language systems (Green & Abutalebi, 2013). The findings from studies in this line of research presented mixed results so far. There has been evidence indicating that bilinguals and trilinguals outperformed monolinguals in inhibitory control or task switching but not showing a substantial difference between bilinguals and trilinguals (e.g., Ma, Wang, & Gao, 2022; Madrazo & Bernardo, 2018; Poarch & Van Hell, 2012; Schwieter & Sunderman, 2011), or conversely, that additional L3 could be associated with a negative effect in inhibitory control and WM functions calling for caution in assuming bilingual and multilingual individuals similar in these abilities (e.g., Guðmundsdóttir & Lesk, 2019). There have also been indications that language switch in trilinguals could be closely related to inhibitory control under particular conditions such as reduced switch costs from and into the dominant language (e.g., Linck, Schwieter, & Sunderman, 2012). Studies investigating L3 effects in relation to other complex WM functions (e.g., Bouffier, Barbu, & Majerus, 2020; Cockcroft, 2022; Espi-Sanchis & Cockcroft, 2019) and LA components (e.g., Ma, Yao, & Zhang, 2018; Thompson, 2013) reported distinctive effects of L3/n experience in different components or tasks depending on various factors such as perceived multilingualism (Thompson, 2013) or socio-economic status and general intelligence (Cockcroft, 2022; Espi-Sanchis & Cockcroft, 2019). The research designs studying the interactions of these two constructs in relation to multilingualism also provided some initial findings such as the enhancing effect of learning two versus one foreign language on LA and WM (T. Huang, Loerts, & Steinkaruss, 2022) and the significant

role of sound recognition ability within LA to predict L2 or L3 group membership (Y. Huang, Wang, & Rao, 2021) among Chinese undergraduate students.

1.2 Purpose of the study

Considering the current state of research regarding WM and LA interaction in L3 learning arising as a new line of study with scarce contexts of empirical evidence yet, the current study aims to examine this interplay in an L3 learning context with students learning additional foreign languages at beginner, intermediate, and upper-intermediate/advanced levels along with an L2-only control group learning no additional languages at an English-medium university in Turkey. Extending the LA and WM research to L3 learning with some representation of L3 comprehension could allow observing the relationships between these two factors as well as their predictive roles in L3 comprehension, probing the effects of enhanced multiple language learning experience beyond a highly proficient L2. Findings can potentially contribute to the discussion on the positions of domain-general abilities underlying WM and domain-specific abilities underlying LA in explaining L3 learning outcomes. Findings could also help understand whether - and if so, how- aspects of WM and LA interact similarly or differently in learning an L3 compared to an L2. To this end, verbal (Operation Span – OSpan) and visuospatial WM span (Rotation Span – RotSpan, and Symmetry Span - SymSpan), phonological short-term memory (digit span) tasks and LLAMA tests version 3 (Meara & Rogers, 2019) are administered to explain L3 comprehension. In this regard, the following research questions are investigated: 1) What is the structural relationship between LA and WM? 2) Do participants learning an L3 differ significantly by their L3 learning experience (i.e., elementary, intermediate, and upper-intermediate/advanced) from those who do not

learn an L3 (L2-only group) in LA and WM factors? 3) Do LA and WM factors significantly explain any additional variance in L3 listening and reading comprehension scores beyond L3 proficiency?

1.3 Significance of the study

The present study aims to address three main areas that need further research and empirical evidence. One aspect is the limited variation in the contexts of studies investigating the roles of LA and WM in the same design for multilingual contexts. The second aspect is the unresolved relationship between LA and WM components, which can still present varying results across different studies. The third point is the lack of skills development research in multilingualism as a growing field in applied linguistics.

The existing studies address multilingualism categorically such as comparing participants based on the number of languages they learn (e.g., T. Huang et al., 2022). The current study addresses multilingualism experience by the number of additional foreign language courses taken (representing additional foreign language learning is an ordinal rather than categorical scale) as well as measuring actual L3 outcome variables operationalized as reading and listening comprehension scores.

The definitions of LA and WM are still an unresolved issue in the foreign language learning field, offering distinctions as well as overlaps in certain components. Findings from different contexts still offer variations across instruments and contexts (e.g., Li, 2016; Wen et al., 2017). This study offers a contribution to the body of empirical evidence from young adults with a different language background that has not been particularly investigated before. Additionally, the study employs

the updated LLAMA version 3 of the widely-used measure of LA along with verbal and visuospatial complex span tasks to measure WM to obtain a more reliable representation of WM.

Multilingualism research is an emergent area of research in applied linguistics. The initial studies tended to focus on particular aspects of language such as morphosyntax, phonology, or pragmatics from a cross-linguistic influence and transfer perspective rather than the development of the language skills (e.g., Amaro, 2017; Bardel & Falk, 2007; Falk & Bardel, 2010). This study particularly presents L3 output represented as reading and listening comprehension. More theoretical approaches such as the dynamic model of multilingualism (Herdina & Jessner, 2002) also introduced a larger perspective of the interaction between languages within a system without a clear emphasis on particular cognitive variables such as LA and WM. This study aims at addressing specific cognitive abilities with relation to multilingual experience and L3 outcome, potentially providing some insight into theoretical as well as practical approaches in multilingualism.

The following section is devoted to the review of literature divided into three main parts: 1) theory of LA and its relation to L2 learning, 2) theory of WM and its relation to L2 learning, 3) interactions between LA and WM in L2 learning, and 4) interactions between LA and WM in L3 learning.

CHAPTER 2

LITERATURE REVIEW

Ever since Miyake and Friedman's (1998) claim of "WM as language aptitude", research on separate and interactional effects of LA and WM in explaining L2 performance has so far provided mixed results about the nature of these two cognitive factors, pointing to both overlaps and distinct aspects concerning L2 learning (e.g. Granena, 2013a; Hummel, 2009; Sáfár & Kormos, 2008; Yalçın et al., 2016). Recent research extends this line of findings to multiple language learning contexts presenting mixed findings in relation to the effects of WM and LA in third or more language learning (e.g., Cockcroft, 2022; Ma et al., 2018; Y. Huang et al., 2021). The current study aims to examine the LA and WM interaction in the L3 learning context adding L3 comprehension as an outcome variable. This chapter presents the theoretical background and existing empirical findings on the LA and WM research in L2 learning. Then, it presents arguments about the relationship between the two constructs in detail. Finally, the current literature on the LA and WM research in L3 learning is presented, concluding with the summary and research questions and hypotheses.

2.1 Language aptitude and L2 learning

2.1.1 Definitions and measures of language aptitude

Language aptitude was originally proposed as a factor that could explain individual variations in second/foreign language attainment. The basic assumption of language

aptitude was that there was a set of particular cognitive abilities related to language learning and strength in these abilities facilitated the efficiency and speed of foreign language learning. Carroll's (1962) four-component aptitude model was widely influential in defining aptitude in the earliest times it was proposed as a construct. In this account, language aptitude was associated with four basic abilities: phonetic coding ability (identifying distinct sounds and retaining them to make associations between sounds and symbols), grammatical sensitivity (recognizing grammatical functions of words in sentences), rote learning (learning associations between sounds and meanings and retaining them) and inductive language learning (recognizing rules that govern the structure of sentences). The model proposed that higher performance in these abilities could be related to a higher rate of learning particularly in earlier stages of foreign language learning (Carroll, 1981).

Various test batteries to detect learners with higher speed and efficiency in learning new languages were developed during the 1960s and 1970s such as the MLAT (Modern Language Aptitude Test, Carroll & Sapon, 1959), PLAB (Pimsleur Language Aptitude Battery, Pimsleur, 1966), and DLAB (Defense Language Aptitude Battery, Petersen & Al-Haik, 1976). Among these, the MLAT received considerable attention during the initial years following its development. It consisted of five sub-tests: 1) number learning (learning words in a novel language), 2) phonetic script (recognizing sound-symbol correspondences in a novel orthography), 3) spelling clues (recognizing English words spelt in novel patterns and matching them with synonyms among options), 4) words in sentences (recognizing the grammatical role of one underlined word in a stimulus sentence and implementing it to other sentences), 5) paired associates (learning words in an unfamiliar language based on their L1 counterparts). These five subsections of MLAT were assumed to

represent the cognitive skills described in Carroll's (1962) definition of language aptitude. "Number learning" was assumed to relate to rote learning and inductive language learning abilities as it involved identifying sound differences and retaining the sound-symbol correspondences. "Phonetic script" aimed at phonetic coding ability as it required recognizing distinct sounds and associating them with symbols. "Spelling clues" was associated with phonetic coding ability and L1 vocabulary knowledge while "words in sentences" was intended to measure grammatical sensitivity and "paired associates" to measure rote learning.

In later reflections on the conceptualization of language aptitude, Skehan (2002) evaluated that the impact that MLAT could achieve as a measure of language aptitude in predicting language attainment remained behind the influence of Carroll's (1962) four-component conceptualization of aptitude which was assumed to represent an ability facilitating foreign language learning. One major reason for this was the criticism that the subtests of MLAT may not represent the skills they were intended to measure (Sawyer & Ranta, 2001). For instance, the spelling clues part requires to adopt a novel rule of writing system which is different from the familiar English orthography, creating a task that goes further than simply recognizing a novel coding system or the words in sentences part favouring formal knowledge of grammatical categories rather than a separate particular language-related skill. Likewise, for the rote learning component, in a later review of his conceptualization of aptitude, Carroll (1990) himself admits doubting the nature of the type of memory represented in the paired associates subtest, which did not as often produce significant correlations as phonological coding or grammatical sensitivity components with foreign language achievement.

In the decades following the introduction of MLAT and similar aptitude tests, which were largely shaped in a predominantly audio-lingual period in second language teaching, the value attributed to the speed of learning became less relevant for instructed environments (e.g. Ellis, 2004). A revival of interest in LA research was observed almost 40 years after its first appearance in the Carrollian approach along with the introduction of alternative test batteries to measure language aptitude. This revival focused on exploring and distinguishing the unresolved relationship of language aptitude with memory structures. Alternative aptitude measures such as CANAL-F (Grigorenko, Sternberg, & Ehrman, 2000), LLAMA (Meara, 2005), and Hi-LAB (Linck et al. 2013) were developed.

CANAL-F was developed based on the cognitive theory of foreign language acquisition emphasizing "the role of coping with novelty in such acquisition" (Grigorenko, Sternberg, & Ehrman, 2000, p. 392). The test aimed to measure the assumed abilities as test takers try to learn an artificial language aiming to simulate real language learning context. The sections of the test were meant to operationalise language-specific cognitive abilities with sections on learning meanings of neologisms from context, understanding the meaning of passages, continuous pared-associate learning, sentential inference, and learning language rules. The construct validity study of the test battery indicated that the CANAL-F subtests stand as separate factors from the MLAT and fluid and crystallized general intelligence tests. The criterion validity study investigating the correlations between language instructor ratings for individuals' foreign language performance and their CANAL-F scores also showed an overall significant and moderate correlation (r = .4), except "understanding meaning from context and passages" not significantly correlating with the instructor ratings.

Hi-LAB, on the other hand, aimed to measure language-specific abilities as well as central cognitive abilities under executive functions, and its design is closer to a more general learning approach than previous aptitude tests (Wen et al., 2017). Linck and colleagues (2013) aimed to measure a "high-level aptitude" defined as "a composite of domain-general cognitive abilities and specific perceptual abilities that, together, can support or constrain one's ability to attain high-level proficiency as an adult learner" (p.535). Therefore, the test battery consisted of tasks assumed to measure executive functions, phonological short-term memory (PSTM), associative memory, long-term memory retrieval, implicit learning, processing speed, and auditory perceptual acuity. A validation study conducted with high proficiency L2 speakers indicated that among these components letter and non-word span (measures of PSTM), paired associates (a measure of associative memory), and serial reaction time (a measure of implicit learning) were significant predictors of higher and lower groups on listening and reading score-based groupings, emphasizing similar language-related skills measured in other test batteries rather than more general executive function subtests (Linck et al., 2013). Confined to a particular group of participants with rarely comparable high proficiency levels so far, Hi-LAB still needs further investigation with different language learner groups.

Largely due to the facility in access and software administration, the LLAMA tests (Meara, 2005) have been a more widely used measure of language aptitude. The conceptual approach behind the LLAMA (Meara, 2005) had significant commonalities with the MLAT; however, LLAMA was designed for use with a larger number of language backgrounds as it eliminated dependence on a particular L1 knowledge to adopt the test. The four subtests are LLAMA B (vocabulary learning), LLAMA D (sound recognition), LLAMA E (sound-symbol recognition),

and LLAMA F (grammatical inferencing) providing counterparts for almost all sections of the MLAT. In a relatively recent study aiming at the construct validation of LLAMA, Granena (2013a) investigated the relationships of these four subcomponents with measures of memory and general intelligence with participants from different native language backgrounds. The results of principal component factor analyses indicated that LLAMA B, E and F loaded on one component with a non-verbal general intelligence measure, while LLAMA D (sound recognition) loaded on a separate component with serial reaction time, namely a measure of implicit learning. This distinction between the four LLAMA tests was interpreted as the LLAMA tapping two different aptitude dimensions, pointing to multiple and differentiated aptitude profiles for the same individuals. The coexistence of an implicit learning measure loading onto the same factor as the LLAMA D also introduced the argument of it as an implicit aptitude measure, with this implicit-explicit aptitude distinction receiving considerable attention in following LA research. LLAMA B, E, and F subtests, involving a separate study phase, were discussed to stimulate more attentive explicit processing of the test materials using strategies and problem-solving mechanisms.

Bokander and Bylund (2020) investigated the internal validity of the LLAMA tests via Rasch analysis for item quality, pointing out reservations about its internal structure. They reported that only LLAMA B produced acceptable rates of item discrimination results while the subtests D, E, and F had problematic items, mostly attributed to its two-option response type, increasing the test takers' chances of correct guessing with the tests failing to reflect the actual construct they aim to measure. Most recently, the LLAMA version 3 with a web-based format was released (Meara & Rogers, 2019), introducing essential fixes in the item response

format. The subtests with the two-option response format were all converted to multiple-choice format, making the correct answers harder to find and requiring detailed processing of the items.

2.1.2 Language aptitude in L2 learning research

Three decades after the initial appearance of LA as a concept in foreign language contexts, the revival of research interest focused more intensively on the correlational and predictive roles of some verbal abilities covered in the definitions of LA with L2 learning outcomes. The change in the climate of L2 research in its view of LA could comply with the more theory-driven LA research considering L2 learning to consist of stages measured at points in time rather than the ultimate level of proficiency to predict at earlier times of L2 learning (DeKeyser, 2019). This view of LA continues up to today's enriched L2 research environment.

The revival of research interest goes back to the late 1990s. One significant study that investigated the role of aptitude in an instructed language learning context was Robinson's (1997) design. The learning of two grammatical rules (one rule was incorrectly judged on a grammaticality judgment test and confirmed as complex by 15 ESL teachers while the other was judged as relatively easy) by a group of L2 learners of English with Japanese, Chinese or Korean L1 background was investigated in four instructional contexts: implicit, incidental, rule-search and instructed conditions. After the instructional intervention, the participants' learning of the rules was measured with a grammaticality judgment test, while aptitude was also measured with the paired associates (rote learning) and words in sentences (grammatical sensitivity) sections of the MLAT. During the intervention, the learners

in the implicit condition were told to remember the position of the words in a sentence thinking they were doing a memory task. The incidental group focused on the meaning of the sentences and answered yes/no type comprehension questions. The rule-search group was told to find the grammar rules in the sentences, and the instructed group was first provided with the rules and their explanations before the sentences. The results indicated that performance on either one or both of the rules was significantly correlated with either one or both of grammatical sensitivity and paired associates sections of the MLAT. However, only the incidental group did not show any significant correlations in their performance on the rules and either of the MLAT sections. This finding was taken as evidence that skills that were assumed to represent aptitude became relevant when some attention is on the form. In an incidental learning condition, which would be more similar to a naturalistic language learning environment with more focus on meaning, neither of the aptitude measures yielded a significant correlation with performance on grammaticality judgment. On the other hand, Erlam (2005) tested the correlations of LA components with target structure learning gains in three instructional conditions (i.e., deductive, inductive, and structured input) while measuring LA through language analytic ability and phonemic coding ability in learning direct object pronouns in L2 French. The gain scores before and after the instruction showed that the deductive group who received explicit rule instruction and had the opportunity to produce language had little significant correlations with LA components compared to the other two instructional groups. Erlam interpreted these findings as indicating that individual differences in LA were neutralised with the effect of deductive instruction with an opportunity to produce L2 output. Similarly, Yalçın and Spada (2016) also observed a differential influence of LA components with grammatical difficulty in L2 English in an

instructed context. They observed that grammatical inferencing component of LA was significantly correlated with the knowledge of passive voice, designated as a difficult structure, while rote memory component was correlated with the knowledge of past progressive, which was an easier structure to acquire in L2 English.

Robinson (2002) in a further research design investigated the roles of aptitude and working memory in an incidental language learning context. The findings lend support for the involvement of aptitude in performance on easy rules in the delayed posttest, which was interpreted as a facilitative influence of aptitude for easily learned aspects later. These extensive findings led to an "aptitude complex" approach pointing to a multi-layered structure of aptitude rather than a monolithic construct.

This multi-layered view of LA was reflected in later research probing the structure of LA and its role in L2 learning focused on various aspects of language such as vocabulary and grammar learning, and the development of the four skills (i.e., speaking, reading, writing, and listening) or general L2 proficiency. It has also been investigated for its interactions with age or language learning experience and other individual differences.

The overall or subtest scores from the LLAMA tests and some subsections of the MLAT were frequently used to measure LA aspects in relation to L2 learning outcomes. In a comprehensive meta-analysis study, Li (2015) presented a systematic review of LA research in L2 grammar learning conducted between the publication of the MLAT and May 2013 were included in the analysis. The findings indicated an overall moderate correlation between LA and L2 grammar learning. It was also more likely to observe significant correlations and predictive roles of LA in younger participants and high school contexts than in older participants and university contexts while aptitude-treatment interaction studies displayed opposite results.

As one of the main areas of LA research, the relationships between age effects and LA pointed to some common findings, emphasizing more significant relations with LA for late learners in explicit conditions processing language at a conscious level with attention compared to implicit conditions with lower levels of conscious processing. In a highly influential work in the field, DeKeyser (2000) tested the fundamental difference hypothesis (Bley-Vroman, 1988) which proposed that adult second language acquisition beyond the sensitive critical age differed from child language acquisition in nature with the former relying on domain-general cognitive abilities such as problem-solving while the latter happened using domain-specific language learning mechanisms. DeKeyser tested this hypothesis with adult Hungarian immigrants in the US with the age of arrival ranging between 1 and 40 years. The findings showed a significant negative correlation (r = -.63, p <.001) between the age of arrival and grammaticality judgement accuracy scores in L2 English. The accuracy in grammaticality judgement was lower in adult arrivals than the early arrivals. When divided into higher and lower language analytic ability groups based on the words in sentences section of the MLAT, the grammaticality judgement scores significantly correlated with the language analytic ability in adult arrival learners (older than age 16) but not significantly correlated in child arrival learners (younger than age 16). These findings were interpreted as evidence that adult arrivals learning their L2 possibly beyond the critical age could attain L2 grammar within the early learner range if they had high language analytical ability, which compensated for the effects of critical age. DeKeyser, Alfi-Shabtay, and Ravid (2010) reported similar results with two different groups of L1 Russian-speaking learners of L2 English in the US and L2 Hebrew in Israel. When controlled for the age at the time of testing, there was a decline in the grammaticality judgement scores

after the age of 18 for arrival in the country where the target language was spoken in both the US and Israel contexts. Significant correlations were observed for the adult arrival learners between their grammaticality judgement scores and language analytic ability measured with the verbal sections of the inter-university psychometric entrance test in Russian to overcome L1 bias measuring verbal ability.

To test whether late L2 learners, who might have little/no access to domain-specific language acquisition facilities due to critical age effects, compensate for this lack via higher LA profiles to attain nativelike proficiency, Abrahamsson and Hyltenstam (2008) compared late and early childhood L2 speakers of Swedish that were perceived as nativelike by native speakers of Swedish in their grammaticality judgement and LA scores measured with an early version of the Swansea LAT tests preceding the LLAMA version 2. They found that late learners who were perceived as nativelike scored above the average of the entire sample on the LA measures and significantly differed from early learners. However, both early and late learners mostly performed below the native speaker average on the grammaticality judgement scores in linguistically demanding structures. Additionally, even among the early arrivals, those scoring in the native range in grammar knowledge had higher scores in LA as well, indicated by a significantly strong correlation between LA and L2 grammar knowledge ($r = .70$, $p < .001$).

The position viewing LA as a compensatory effect mainly through language analytic ability in a mostly domain-general explicit language learning was not completely confirmed with more layered micro-level studies finding LA effects in early L2 learners when they differentiate L2 grammar structure by complexity/difficulty or grammar knowledge measures by modality to represent implicit or explicit L2 knowledge (e.g., Granena, 2013b; Granena, 2014). The

different aspects of L2 knowledge also interacted differently by age effects with windows of opportunity closing first by phonology followed by lexis and morphology correlating with LA for late learners (e.g., Granena & Long, 2012).

Regarding the four language skills, L2 speaking received considerable attention in the LA research. Saito (2017) reported grammatical sensitivity, rote learning, and phonemic coding to significantly correlate with fluency, pronunciation, and lexical diversity factors of the oral ability of Japanese college students learning L2 English. More recently, Granena (2019) reported the LLAMA D sound recognition subtest and the available long-term memory subtests of the Hi-LAB constituting a factor labelled as "implicit memory ability" to significantly predict the speed of L2 fluency.

The relationships of LA with general L2 proficiency were also investigated offering a differentiated relationship across the proficiency levels. Artieda and Muñoz (2016) reported that the overall scores from the LLAMA tests had a moderately significant correlation with overall L2 English proficiency of Catalan-Spanish adult bilinguals (r = .39). However, LLAMA D, the test of sound recognition significantly explained 16% of the variance in L2 proficiency only in the beginner group (N =52), while LLAMA F, the grammatical inferencing test explained 8% of variance only in the intermediate group (N = 88).

Evaluating the definitions and measurements of LA as a construct and the overall findings of empirical research presenting data using these instruments, Skehan (2016) proposed that different aspects of what was defined as LA could be relevant at different stages of L2 learning and that the majority of the tests available correspond to the initial stages of L2 learning, leaving space for different measurements of LA for the later and more proficient stages. The main assertion of

Skehan's stages model defined L2 learning as a process starting with building and organising L2 knowledge at earlier stages, and then moving on to controlling this knowledge in production and use at later and more advanced stages. The stages of L2 leaning were ordered from bottom to top as input processing, noticing, pattern identification, complexification, handling feedback, error avoidance, automatisation, creating a repertoire, and lexicalising (Skehan, 2016; 2019). Skehan (2019) further categorised the first two stages to focus on handling sound, while identifying patterns through feedback on handling pattern. The stages of error avoidance and onwards were categorised as automatisation-proceduralising. Skehan argued that the existing views, particularly the Carrollian model of LA and the measures based on them correspond to the first two macro-categories, namely handling sound and pattern. What remained largely missing in theory and measurement were automatizing-proceduralisation stages also involving implicit learning. Skehan (2019) pointed out that although some attempts such as CANAL-F involving language learning as a continuum aiming at implicit learning processes or the LLAMA D categorised to fall closer to implicit learning abilities, there was still a need for a systematised measurement with a consistent underlying theory of implicit L2 knowledge.

At this point of locating the LA as an important construct researched in the L2 learning field, it becomes indispensable to discuss the involvement of WM, involving individual differences in abilities to store and process information, and management of online cognitive processes, in describing LA. The following section of this chapter first presents a general description of WM and then its role in L2 learning research. Then, a review of research investigating the LA and WM in common designs in L2 and L3/multilingual learning contexts is presented.

2.2 Working memory and L2 learning

2.2.1 Definitions and measures of working memory

The basic definition of working memory as a limited cognitive capacity of storage and processing first proposed by Baddeley and Hitch (1974) describes the term in the most common and simple way it is addressed today. The latest version of the working memory model of Baddeley (2000) involves four sub-components: 1) phonological loop 2) visuospatial sketchpad 3) central executive 4) episodic buffer (Baddeley, 2000). The phonological loop and the visuospatial sketchpad, which are also called slave systems, are the two storage components and their functions are regulated and coordinated by the central executive. The phonological loop is defined as the acoustic-verbal sub-system which is responsible for temporarily storing and manipulating auditory information within a limited capacity range. The loop is proposed to contain two subcomponents in itself: temporary storage and articulatory rehearsal. The temporary storage works together with the articulatory rehearsal in that the rehearsal system makes it possible to keep the limited information refreshed through subvocal repetition of the items. In cases when the articulatory is suppressed through, for instance having an individual continuously say an irrelevant word “the” while trying to remember lists of digits, the performance in remembering the items is observed to be impaired. This is taken as evidence that the articulatory rehearsal facilitates registering of information to the phonological store.

The visuospatial sketchpad is the subcomponent compatible with visual-spatial information and is proposed to be divided into “visual, spatial, and kinaesthetic components” (Baddeley, 2000, p.2). The central executive is assumed to be involved in regulating the simultaneous work of storage and processing by controlling the slave systems. It is also proposed that communication among the

slave systems is possible through the central executive. The episodic buffer was added to the initial model that consisted of the central executive and the two subsystems. In his revision of the three-component model in 2000, Baddeley discussed the compatibility of this model with some empirical evidence showing that while articulatory suppression is applied and the items to be recalled are presented visually, the performance is not dramatically affected as would be expected when the items were presented aurally. In addition, individuals with poor PSTM recall more items when they are presented visually. This observation created a space in the model which points to a workspace where both the auditory and visual information is processed and computed rather than the separated loop and the sketchpad. This would suggest that such a workspace could also make it possible to connect the temporary items registered to the short-term memory temporarily with items from the long-term memory through meaningful networks to make the presented items last longer, which were attributed to the episodic buffer.

The executive attention model (Engle, 2002; Engle & Kane, 2004; Engle, Tuholski, Laughlin, & Conway, 1999) assumed working memory as an attentional function that is different from STM. In their view, working memory is related to controlled attention capacity rather than a multi-levelled system involving STM-related subcomponents. Their main argument is that individuals with high WM span (measured through span tasks which will be discussed in detail later) are better at resisting interference by irrelevant stimuli while performing a cognitive task compared to individuals with low WM span. Therefore, they state that "...WM capacity, the construct measured by WM span tasks, reflects the general capability to maintain the information, such as task goals, in a highly active state. Although the need for such active maintenance will be minimal in many contexts, it will be

particularly important under conditions of interference" (Kane, Bleckley, Conway, & Engle, 2001, p. 170).

These differences in the conceptualization of WM were accounted as "unitary and non-unitary camps of working memory" models by Shah and Miyake (1996) pointing to the componential and more compact systems. However, in terms of functions and abilities attributed to WM as a construct, these models are not incompatible with one another (Dehn, 2008).

In earlier times of WM research, simple memory span tests that require the individuals to recall as many items as possible from auditorily or visually presented lists of digits, letters or non-words were more commonly used. In these simple span tests, the maximum number of correctly remembered items is generally accepted as the score and indication of the span of the individuals. However, considering the functions proposed by the theoretical models, measures of WMC involving the simultaneous processing and storage ability were required to investigate the relationship of WMC with more complex cognitive functions such as language comprehension, performing mathematical operations or reasoning. It was also observed that simple span tests were not significantly related to complex functions such as language comprehension (e.g., Daneman & Merikle, 1996).

Rather than measuring the more complex functions attributed to WM, these simple span tests were more appropriately measured short-term memory emphasizing the storage capacity. Reading span tests (RST) became widely used to measure WM span separately from simple span tests which were agreed to tap short-term memory, which did not reflect the processing aspect of the WM as a construct. The version of RSTs developed by Daneman and Carpenter (1980) involved reading sets of sentences aloud and keeping the last word of every sentence to be recalled at the end

of a set. The number of correctly recalled words was recorded as the WM span of the individual. The RST scores observed on these tests yielded significant correlations with individuals' reading comprehension in language reading comprehension (e.g. Daneman & Carpenter, 1980; Friedman & Miyake, 2004).

However, Waters and Caplan (1996) argued that the significant correlations observed between Daneman and Carpenter's (1980) RST and reading comprehension could be due to the overlap of functions performed in RST and reading comprehension tasks rather than a capacity that shapes the two cognitive tasks. They criticized Daneman and Carpenter (1980) in that simply reading aloud or hearing sentences (in listening span versions) might not necessarily direct the individuals to process the sentences at deep levels as would be observed with online sentence processing. Alternative span tasks involving grammaticality judgment (Turner & Engle, 1989) were also criticized for the possibility that the processing tasks might not be essentially triggering a sentence processing function for semantic search or syntactic computations. Therefore, they suggested a version of a complex RSTs that involved sentence processing tasks (acceptability judgment) in addition to a recall task and the calculation of scores included both processing (as measured in time taken to react to an item) and accuracy (the number of correctly recalled items). Their observation indicated the processing component in the span tasks had a significant determinant role in predicting language comprehension (Waters & Caplan, 1996, p. 76).

Alternative complex WM span tests involved non-linguistic secondary tasks in the verbal domain. Turner and Engle (1986; 1989) aimed to test whether the significant correlations obtained in Daneman and Carpenter's (1980) reading span test and reading comprehension would still be observed in tasks where the secondary

task is not of a similar nature as reading. For this purpose, they replaced the sentence reading task with confirming arithmetic operations before the recall items were presented. The task named operation span (OSpan) was also significantly related to reading comprehension along with an RST in which the secondary task is also reading sentences. This finding confirmed that the secondary task which involves the processing aspect of WMC did not have to be similar to a complex WM span task.

Conceptualizing WM as a construct involving multiple complex tasks using verbal and visual-spatial domains, Kane et al. 2004 conducted a validation study on WM as a domain-general capacity involving verbal and spatial span tasks while establishing more distinct verbal and visuospatial STM components. The design included three complex WM tests (with primary and secondary tasks as proposed in Waters and Caplan's 1996 study) as well as simple STM span tests on verbal and spatial domains.

Complex verbal WM tests (i.e. operation span -OSpan, reading span - ReadSpan, and counting span - CouSpan) involved verbal stimuli (i.e. either digits or words) in both primary (storage) and secondary (processing) tasks. Complex visual WM span tasks (i.e. rotation span - RotaSpan, symmetry span - SymmSpan, and navigation span - NavSpan) involved visual/spatial stimuli in both primary (storage) and secondary (processing) tasks. For instance, RotaSpan required to mentally rotate mirror-reversed or normal letters of G, F and R presented as rotated at various degrees and tell whether the letters were mirror-reversed or not while recall stimuli were short or long arrows heading in different directions.

Simple WM tasks measuring STM and involving only storage tasks were also verbal (i.e. word span, letter span, digit span) and spatial (i.e. arrow span, matrix span, and ball span). A confirmatory factor analysis indicated that the best model fit

was achieved in the model in which complex verbal and spatial WM tasks loaded on WM as one factor while verbal and spatial STM tasks loaded separately on verbal STM and spatial STM factors (Kane et al., 2004, p.200). The relationships of WM as a latent variable were further examined with measures of general fluid intelligence as well as spatial and verbal reasoning. Path analysis indicated that WM was a significant predictor of general fluid intelligence (r = .64) while it was not true for spatial and verbal reasoning measures. Verbal and spatial STM measures were better predictors of these domain-specific latent variables.

In general, these findings provided support for the view that WMC is a more domain-general ability representing functions of simultaneous processing and storage attributed to executive functions while simple storage functions related to STM could be more domain-specific in terms of the modality of stimulus being recalled.

2.2.2 Working memory in L2 learning research

The role of WM as measured through simple and complex WM span tests in L2 learning was investigated in various research contexts. An overall view of the accumulation of empirical findings so far pointed to differentiating results with different L2 learning outcomes such as general proficiency, vocabulary or grammar learning, or receptive and productive language skills (Juffs & Harrington, 2011). However; Linck, Osthus, Koeth, and Bunting (2014) reported a systematic review of WM in L2 research presenting an overall robust positive relationship between L2 outcomes and WM in verbal rather than non-verbal and processing rather than storage components.

Initial studies in WMC and L2 learning indicated significant relationships of RSTs performed in L2 with Daneman and Carpenter's (1980) design with L2 reading comprehension (e.g. Osaka & Osaka, 1992; Osaka, Osaka & Groner, 1993), and non-significant correlations of simple span tests with L2 reading comprehension (e.g. Harrington and Sawyer, 1992).

Some studies investigated how different components of WM could be related to aspects of novel L2 learning. For instance, Williams and Lovatt (2003) reported that performance on a rule generalization task in an unfamiliar language was related to measures of PSTM, vocabulary learning and memory for the target rule. The more consistent correlations of vocabulary learning measure (recalling words in the novel language with L1 equivalents) with more sections of the rule learning task was interpreted as a potential mediating role of vocabulary learning between grammar learning and PSTM due to significant correlations of PSTM and vocabulary learning tasks. Pointing to the role of PSTM in L2 vocabulary learning, Martin and Ellis (2013) also reported vocabulary learning as a mediating factor for the relationship between PSTM (non-word recognition and non-word repetition), WM (listening span) and grammar learning in an artificial language learning context.

The interaction of L2 proficiency and WM was also investigated in several studies. In a cross-sectional study with a group of L1 Finnish L2 English, Service et al. (2002) compared WM spans performed in L1 and L2. The task was in the form of a complex listening span which requires choosing the correct picture describing the meaning of the sentences heard while recalling the last words of the sentences. The participants were divided into a lower and higher L2 proficiency group. The initial results indicated only the lower proficiency group had significantly higher span scores (more correctly recalled words) in the L1 than the L2, while the more

proficiency group did not have a significant difference in the two versions of the WM span. In an additional experiment, measures of PSTM (simple word span and colour naming) indicated no significant difference between the two groups. Therefore, the difference observed between higher and lower proficiency groups in WM tasks involving sentence comprehension and the comparable L1 and L2 performance of the more proficient group was interpreted as evidence for an interaction between WMC and L2 proficiency. At higher L2 proficiency levels, there could be less load on WM and performing similarly to L1.

Service et al.'s (2002) conclusion about the interplay between proficiency and WM received some sort of an extension in Serafini and Sanz (2016), examining this issue in a longitudinal cross-sectional design. Three groups of university students in an L2 Spanish learning context in three proficiency levels (beginning, intermediate, advanced) performed on simple digit span (phonological WM), OSpan (complex verbal WM) and L2 Spanish grammar knowledge tests. It was observed that both measures of WM were related to L2 grammar performance and this connection disappeared in the advanced group. The researcher interpreted their findings that phonological WM and complex cognitive functions of WM were related to L2 learning only in the early stages of proficiency. As L2 proficiency increased, the effects of these factors weakened.

The effects of WM across L2 proficiency levels were investigated in simple and complex WM measures in Mitchell, Jarvis, O'Malley, and Konstantinova's (2015) study. The data from 36 Chinese-speaking undergraduate learners of L2 English at a US university indicated that the L1 digit span had a weak and marginally significant correlation with the TOEFL iBT scores only for the beginner group. On

the other hand, L2 digit span and OSpan scores all produced non-significant correlations with iBT scores.

The processing and storage components of WM with L2 English knowledge and reading comprehension were also differentially examined in Çeçen and Erçetin's study (2016) with L1 Turkish-speaking young adults with extensive L2 English instruction background. They observed that the processing and storage components of complex reading and operation span tasks were loaded onto separate factors. The untimed grammaticality judgement for explicit and elicited oral imitation for implicitL2 knowledge loaded together on a different factor. These three factors were regressed on L2 reading comprehension. The processing and L2 linguistic knowledge factors were significant predictors of L2 reading comprehension. These findings highlighted that it was the processing aspect of WM that was related more directly to L2 reading comprehension rather than their linguistic material (i.e., recalling words or letters) but rather the non-linguistic processing (i.e., arithmetic operations or semantic acceptability judgement). Linguistic knowledge played its role as a separate factor from the WM measures operationalised as oral imitation and grammaticality judgement scores.

Based on the findings of WM research in L2 learning studies and the prominent definitions of WM, Wen (2015) proposed the phonological/executive model describing the WM functions specified for L2 learning. Reflecting Baddeley's three-partite model and the North-American approach of executive functions, this integrated framework of WM in L2 learning encompassed the verbal components of storage as the phonological aspect reflecting the domain-specific aspects (i.e., phonological short-term storage and articulatory rehearsal) and the domain-general aspect of executive functions (i.e., task switching, inhibition, and updating). Wen

also included long-term memory as an inherent part of this structure taking part in the declarative/procedural memory in a bidirectional relation with the phonological and executive components. The phonological WM was positioned to underlie the building of lexis, formulaic sentences, and morphosyntactic structures. The executive memory was assumed to relate to deeper language processing such as ambiguity resolution, pronoun detection, subject-verb agreement, or resolving the antecedents in relative clauses. In his following work, Wen (2019) proposed that the "WM as language aptitude" postulation might be explained with this model emphasizing the language-related domain-specific aspects of the WM models.

The overall review of WM research in L2 learning demonstrates the roots of the intersections bringing LA and WM as two cognitive constructs still posing areas to be systematically distinguished and defined. The multi-layered models of both LA and WM propose areas of L2 learning still call for empirical evidence to reflect on questions such as whether L2 learning particularly in instructed contexts induces domain-general learning mechanisms (e.g., DeKeyser, 2019) or still involves domain-specific language nature in essence (e.g., Skehan, 2019) with the automatisation at later stages. The following section presents research involving LA and WM measures in common designs to explore their interactions or probe the roles of two constructs related to L2 learning outcomes.

2.3 Language aptitude and working memory interaction in L2 learning

Although aptitude gained considerable acceptance as a potentially significant individual factor in L2 learning, its inherent structure and connections to memory functions were still under scrutiny for a long time (Ellis, 2004; Skehan, 2002). In the

light of a study on the roles of WM measured in L1 and L2 in acquiring cue preferences and syntactic comprehension, Miyake and Friedman (1998) proposed that it could be WM functions that involved the abilities which were called language aptitude by making the "WM as language aptitude" claim. Their study involved 59 native Japanese-speaking learners of English in performing four tasks: a WM listening span task in L1 Japanese (listening to sentences and recalling the first word in every sentence), a listening span task in L2 English (same as the L1 WM test with the last word of every sentence recalled), cue preference (identifying agents in sentences with word order, animacy or agreement being the correct cue for role assignment), and a syntactic comprehension task (listening to English sentences and drawing a diagram to depict the thematic roles of the nouns). All bivariate correlations between the four types of measures were significant with moderate magnitudes (ranging from .30 to .58). A path diagram was produced with these measures to explore directional relationships, which indicated that the L1 WM span was related to the L2 WM span with a single-headed arrow implying a causal relation from L1 WM to L2 WM. L2 WM, on the other hand, was connected to both cue preference and syntactic comprehension with single-headed arrows indicating the explanatory power of L2 WM on cue preference and syntactic comprehension. Finally, one single-headed arrow from cue preference to syntactic comprehension also implied a predictive role of the former to the latter component. Among these variables, cue preference, which is not explicitly taught in instructional environments, could be accepted as a grammatical aspect that could be supported by grammatical sensitivity as part of language aptitude and described as recognizing underlying rules governing sentence structures. This ability facilitated syntactic comprehension, both of which were also supported by L2 WM. Miyake and

Friedman interpreted these findings as evidence that WM resources came from a shared pool as they were related and this WM capacity supports certain language-specific skills in L2 learning.

Studies implementing WM and aptitude measures in relation to foreign language learning present varying findings. Hummel (2009) investigated aptitude, PSTM and L2 proficiency with highly proficient L2 English learners. Three of the subtests of MLAT (paired associates, spelling clues, and words in sentences), an English proficiency test on vocabulary, grammar and reading sections, and a non-word repetition for PSTM were administered. Bivariate correlations indicated no significant relationships between PSTM and any of the subcomponents of aptitude. Exploratory factor analysis showed that the subtests of aptitude, proficiency and PSTM loaded on separate factors, pointing to a distinction between aptitude subtests (of grammatical sensitivity, rote learning and phonetic coding) and PSTM. Multiple regression analysis indicated words in sentences subtest of aptitude representing language analytic ability and PSTM as significant predictors of L2 proficiency with partial correlations of .14 and .10 respectively.

In a longitudinal study, Sáfár and Kormos (2008) observed the effects of WM, PSTM and aptitude on the L2 English development of Hungarian learners in a focus-on-form instruction context. The Hungarian version of the MLAT was implemented with its five subtests along with the backward digit span test for WM and non-word span tests as the measure of PSTM. The learners' English proficiency was also measured on four skills and general use of English. The pre- and post-administrations of the aptitude tests indicated some significant moderate correlations (ranging between .32 and .44) of several subtests on different proficiencies (grammatical sensitivity with reading and writing only in the pretest, vocabulary

learning with reading only in the posttest) while total proficiency was significantly correlated with overall aptitude scores both in pretests (r = .34) and posttests (r = .36). None of the subtests or the overall scores of aptitude in pre and posttests was significantly correlated with PSTM. However, significant moderate correlations were observed between overall aptitude scores with WM measure on both pre- and posttest and only the language analysis subtest was significantly correlated with WM (r = .33). Sáfár and Kormos interpreted their findings as an indication that while aptitude and WM interact in certain aspects such as language analytic ability, there could also be distinctions in the abilities they might be related to.

The tendency of significant but moderate correlations among measures of WM and LA subtests in the L2 learning context was supported in Yalçın et al. (2016) study as well. The study investigated the relationships of WM and LA with L1 Turkish-speaking university students with advanced-level L2 English. The measures of WM were two RSTs given in L1 and L2, OSpan in L1, and the four subtests of the LLAMA. Their results also showed that only LLAMA F (grammatical inferencing) was significantly related to RSTs in L1 (r = .41) and L2 (r = .32) but not with the OSpan score.

The studies investigating the relationships of LA, WM, and PSTM measures often indicated significant but moderate correlations between language analytic/grammatical inferencing subtests and complex WM measures, supporting the LA and WM as related but not interchangeable cognitive abilities with relation to L2 learning (Robinson, 2002). The most recent accounts also tend to agree that LA and WM could be consisting of several subcomponents that could overlap at various language-supporting interfaces but not potentially build inherently interchangeable constructs (Skehan, 2016; Wen et al., 2017).

The interpretations of a multi-componential structure of LA subtests were proposed in Granena's (2013) influential study on the validity of the LLAMA, offering an explicit-implicit categorization among the subtests. The study reported principal component analysis findings indicating LLAMA D loading on the same factor as attention control and implicit learning as opposed to the other three subtests of B, E, and F, loading on a separate factor with a measure of non-verbal general intelligence. Measures of WM and processing speed remained separate factors in this study. Granena interpreted this observation as evidence that LLAMA B, E and F sections tap a language analytical ability as these sections required watching for regularities and consciously focusing on this with a self-study session allowed before the test. LLAMA D, on the other hand, did not include a previous study session and required instantly recognizing sounds, implying a more implicit perceptual ability.

This factorial structure was observed in different studies. Xiang and colleagues (2012) found with young adult German or Dutch-speaking learners of L2 English that LLAMA B, E, and F loaded on the same factor as readings span as a measure of verbal WM, while LLAMA D negatively loaded with the general cognitive ability. Spatial WM measure remained a separate factor. Similarly, Yalçın et al. (2016) reported findings showing LLAMA B and F loading on the same factor separated from LLAMA D and E. LLAMA F also significantly correlated with both L1 and L2 reading span scores, in a similar line with Xiang et al. (2012) findings confirming the LLAMA B and F relationship with verbal WM.

These interactions between different subtests of LA and WM showed variations across studies in different contexts. Investigating the predictive roles of WM and language analytic ability in explaining generalisation on an artificial language material with Japanese-speaking young adults, Dardon and Jeong (2021)

found that only the OSpan was a significant predictor. In a study with English-speaking learners of L2 Chinese at an advanced level, Winke (2013) reported that rote memory and grammatical sensitivity were the strongest tests of the MLAT to load onto the LA latent variable. However, LA was not a significant predictor of L2 outcomes among phonological WM, motivation, and strategy use. Only strategy use was a small predictor of L2 reading.

From a wider theoretical perspective, it is important to clarify this interactive picture regarding the two constructs termed LA and WM due to the memory and knowledge representations they might be associated with to understand how and when domain-specific or domain-general cognitive processes take place in L2 acquisition. Mostly defined by particularly unitary accounts as a domain-general cognitive resource encompassing various processing and storage functions, WM has been observed to be involved in non-linguistic along with linguistic functions (e.g., Redick et al., 2012). LA, on the other hand, has been defined as a cognitive resource specialised in processing language and linguistic material, thus as a set of domain-specific abilities. Whether one of them lies within the structure of the other as a special area or the two constructs are fundamentally different is a question that has not been fully answered in the L2 learning field. While the majority of findings did not report essentially strong correlations, implying the two constructs did not represent the same latent variable (e.g., Li, 2016), there is also not a consistent map of the abilities composing these two constructs in domain-general and domain-specific areas, leaving questions about whether (and how) LA and WM as key cognitive individual differences are involved in building L2 knowledge within the memory systems. The theoretical approaches discuss how L2 learning occur with relation to the level of consciousness and attention to language input with different

positions under the “automatisation” issue, referring to explicit L2 knowledge built through conscious processing could becoming proceduralised, namely implicit, requiring less attention and executed below consciousness level in time through continuous use (e.g., DeKeyser, 1997). While some approaches leave place for both directly implicit and explicit processes to occur while building L2 knowledge (e.g., Ullman, 2005), some argued that L2 learning essentially happened through implicit processes below the level of consciousness along with explicit processes (e.g., Paradis, 2009). More recently, DeKeyser (2019) emphasized the importance of understanding and measuring language-specific implicit L2 learning processes as the majority of what LA research so far seemed to capture explicit L2 learning mechanisms above the level of consciousness with language analytic ability. Skehan (2019), however, highlighted the essential place of language-specific aspects even in views of L2 learning as a domain-general process:

> Interestingly, both of these positions (a unified approach and a critical period approach) can be argued to mean that language, in the second language, post-critical period case, is not special, and general learning (and therefore aptitudes based on general learning principles) are what we need to focus on. (…) I want to argue that, despite these possibilities, language is special, even while assuming the existence of a critical period and that a concern to embed language tasks and material within foreign language aptitude tests is indispensable. (Skehan, 2019, pp. 56-57)

It remains a significant research question to understand the possible interactions of the LA and WM within domain-specific and domain-general areas of the memory system representing L2 abilities.

Wen and Skehan (2021) evaluated this interactive relation between the LA and WM integrating their recent approaches to these two constructs. Skehan’s (2016, 2019) stages approach to defining LA highlighted the potential roles of the traditional LA subtests in explaining the initial stages of handling sound and pattern. Wen’s (2015, 2019) phonological/executive model of WM focused on the verbal and

phonological components of the WM models emphasizing the roles of phonological and executive components of WM involved in acquiring forms or processing/interpreting complex meanings from structures. They proposed that these two models can be compatible, requiring more attention for the higher levels of proficiency beyond the existing measures of LA.

2.4 Language aptitude and working memory interaction in L3 learning

The body of research in L3 learning exists as a very young and promising field growing more rapidly in the last decade. The research questions that have been investigated so far include such topics as cross-linguistic transfer in various language domains such as syntax (e.g., Bardel & Falk, 2007; Falk & Bardel, 2010), morphosyntax (e.g. Amaro, 2017), or language skills development in academic contexts (e.g., Lasagabaster, 2010).

Very recently, L3 learning received attention concerning cognitive factors such as inhibitory control, task switching or verbal fluency through the research on cognitive advantages of bilingualism. Findings showing superior performance of bilinguals on tasks measuring inhibitory control, task switching or updating (e.g. Bialystok, 2009; 2011) compared to their monolingual counterparts inspired trilingual studies investigating whether an L3 would enhance these effects attributed to bilingualism. Findings in these studies mainly repeated bilingual studies in displaying bilingual or trilingual advantages over monolinguals but not necessarily trilingual advantages over bilinguals (e.g. Schwieter & Sunderman, 2011; Poarch & Van Hell, 2012).

One pioneering study was reported by van den Noort, Bosch and Hugdahl (2006) on WM and foreign language proficiency in an L3 learning context. The study investigated whether WMC is language-specific in languages of differing proficiency levels in third language learners. The participants were L1 speakers of Dutch, fluent L2 speakers of German and beginner learners of L3 Norwegian. Two complex (RSTs in the three languages, and a letter-number ordering span) and a simple (digit-span with one of three languages used for response in separate administrations) WM span tests measuring accuracy and reaction times were administered. The results indicated significantly higher (more accurate responses) and faster (smaller reaction times) performance in L1 compared to L2 and L3 in the complex span tests (i.e. RST and letter-number ordering). This finding was further observed for L2 compared to L3 in RST scores but not for letter-number ordering scores, which did not yield significantly higher or faster performance in L2 compared to L3. The simple digit span task, on the other hand, was reported to indicate significantly better performance only in L1 compared to L3, however no significant differences between L1 and L2 or L2 and L3. No significant correlations were found between simple and complex WM span tasks. These findings were interpreted as evidence for separate resources related to performance in languages of different proficiencies. Whether WM span was language-specific as languages measured in earlier studies belonged to different language families was further tested by comparing L1 speakers of the three languages (i.e. L1 Dutch, L1 German and L1 Norwegian speakers) on all WM span tasks finding no significant differences among the three L1 language speakers and pointing no evidence for language-specificity of WM span tasks

Very recent studies investigating WM interactions with multiple language experiences presented varying results. Cockcroft, Wigdorowitz, and Liversage (2019) implemented processing and storage measures of WM in two modalities (i.e., verbal and visuospatial sections of automated WM assessment battery by Alloway, 2007), and verbal and non-verbal general intellectual ability (WAIS-III), comparing multilingual L2 English learners speaking three to five South African local languages as L1 (N = 39) to their monolingual native English-speaking counterparts (N = 39) all studying at a South African university. They also controlled for socioeconomic status (SES) due to its potential effect on linguistic and intellectual abilities in the particular context. They observed that the multilingual group had significantly lower scores in SES and verbal intelligence. When controlled for these two variables, the multilingual group significantly outperformed the monolingual group in both storage and processing components of WM in both modalities. Cockcroft (2022) investigated the issue further for any distinguishing effects of the number of multiple languages with a comparable sample (N = 162) in the South African context. Implementing the same set of instruments as the former study, this design revealed that the number of spoken languages was negatively associated with the visuospatial storage and processing components of WM. However, Espi-Sanchis and Cockcroft's (2022) study did not completely confirm these findings when the multilingualism effect was measured for balanced proficiency across all the languages using the self-reporting Language Experience and Proficiency Questionnaire (LEAP-Q) with a comparable sample (N = 189) in the same context. This study revealed that between-language balance across an individual’s languages was a significant predictor of the verbal processing component of WM.

Bouffier et al. (2020) reported findings from trilingual speakers of Luxembourgish, German, and French completing auditory-verbal WM and verbal and visuospatial attentional tasks. Measured with an auditory immediate serial recall of words and non-words in the L1, the auditory-verbal WM was significantly associated with non-native language proficiency measured with receptive and productive vocabulary knowledge.

The interactions of LA with the multiple language experience also received attention. In their overall evaluation of the LLAMA tests with findings from various contexts, Rogers and colleagues (2017) reported that instructed L2 learners outperformed the monolingual group in LLAMA B (rote learning) and F (grammatical inferencing), while no significant differences were found in LLAMA D (sound recognition) and E (sound-symbol correspondence). Ma et al. (2018) compared young adult Chinse-speaking learners of L2 English learning additional foreign languages (i.e., Japanese or French) to those who did not learn any additional L3 beyond their L2 English in their LLAMA subtest scores. They found that the L2 and L3 groups significantly differed in LLAMA B and E scores, while no significant differences were observed in LLAMA D and F tests. Additionally, the L3 Japanese group marginally outperformed the L3 French group in LLAMA B, interpreted as a facilitative effect of the typological similarity between the L3 and L1. On a different LA measure, Thompson (2013) also reported a positive relationship between the previous language learning experience and overall CANAL-F scores.

Two recent studies investigating the interaction of LA and WM components in L3 contexts also provided some initial findings in the field. In a pre and post-test quasi-experimental design, T. Huang et al. (2022) the effect of learning L2 English only compared to learning an additional L3 (additional Japanese or Russian) with

young adult undergraduate students in China. The LLAMA tests and the OSpan task were implemented before and after one academic year of L2 and L3 learning. Their results showed that while there were improvements in LLAMA subtests and Ospan from pre- to post-test as repeated measures, the L2-only and L3 groups significantly differed in their improvement intensity in WM.

In a cross-sectional design, Y. Huang et al. (2021) compared accomplished bilingual young adults with counterparts learning additional L3 Japanese at a university in China in LLAMA tests and nonverbal WM task (fractal 2-back task). They reported that the L3 group significantly outperformed the L2-only group in LLAMA D (sound recognition) and the nonverbal WM task. A linear discriminant analysis also indicted LLAMA D as the only significant predictor of L2 or L3 group membership.

Considering the scarcity of studies representing only particular participant contexts, investigating the roles of WM and LA inL3 learning in diverse samples can provide valuable insight into the understanding of these cognitive individual differences potentially generalizable to a population. First of all, implementing L3 learning aspect allows observing a developing language in individuals who have already achieved considerable proficiency in L2 with substantial active communicative use in real life. Secondly, observing this process in relation to WM and LA could help us understand whether, and if so, how these factors are extensively (even further than an L2) influential in learning languages. If there are differences in the patterns of these constructs' relationships with L3 from those already observed in L2, then one could draw implications on how additional languages develop.

2.4 Summary and the research questions

Ever since Miyake and Friedman's (1998) claim of "WM as language aptitude", research on separate and interactional effects of WM and LA in explaining L2 performance has so far provided mixed results about the nature of these two cognitive factors, pointing to both overlaps and distinct aspects concerning language learning (e.g. Hummel, 2009; Granena, 2013; Sáfár and Kormos, 2008; Yalçın et al., 2016). The current study aims to further examine this interaction introducing the L3 learning context as a novelty to the field.

The following research questions and hypotheses are proposed:

i. What is the structural relationship between LA and WM measures?

ii. Do participants learning an L3 at different proficiency levels (i.e., elementary, intermediate, and upper-intermediate/advanced) differ significantly from those who do not learn an L3 (L2-only group) in LA and WM factors?

iii. Do LA and WM factors significantly explain any additional variance in L3 listening and reading comprehension scores beyond L3 experience?

The first research question is mainly an exploratory question aiming to observe whether the previous findings could be confirmed in the specific context of the study. However, a common finding from previous research is a construct-level separation between the measures of LA and WM (e.g., Granena, 2013, Xiang et al., 2012; Yalçın et al., 2016). Therefore, the LLAMA tests and the complex WM tests are expected to stand out as separate factors (Hypothesis 1). Previous research also reports a multi-faceted structure in the LLAMA tests, particularly differentiating LLAMA D from the other three. These subtests have also been found to display

different relations with different outcome and predictor variables (e.g., Granena, 2013; Yalçın et al., 2016). However, no particular directionalities or factor distributions are predicted due to the variability of these findings across different contexts. The verbal and visuospatial WM scores are also predicted to load onto separate factors as due to the type of input material (e.g., Foster et al., 2015) (Hypothesis 2).

The second research question probes the relationship of foreign learning experience with LA and WM factors. Earlier research on L2 proficiency reports some correlations with LA (e.g., Artieda & Muñoz, 2013) and WM (e.g., Serafini & Sanz, 2016) with language learning experience. More recent research extending the question to the L3 learning context also reports varying effects of L3 proficiency on LA (e.g., Ma, et al., 2018; Thompson, 2013) and WM (e.g., Bouffier et al., 2020; Cockcroft, 2022; Cockcroft et al., 2019). Therefore, some significant effect of L3 proficiency on LA and WM factors is expected with no specific hypothesis about the direction of the effect (Hypothesis 3).

The third research question investigates the predictive roles of LA and WM factors in explaining L3 listening and reading comprehension. The reported relationships of LA and WM with the two productive skills are varied (e.g., Alptekin & Erçetin, 2011; Kara-Duman, Yalçın, & Erçetin, 2021; Leeser, 2007; Sparks, Humbach, Patton, & Ganschow, 2011). Therefore, some predictive role of LA and WM factors is assumed for L3 listening and reading comprehension scores (Hypothesis 4).

CHAPTER 3

METHOD

The current study aims to examine the LA and WM interaction in an L3 learning context. The participants are L1 Turkish-speaking undergraduate students at an English-medium instruction university in Turkey majoring in an English language teaching program with extensive L2 English experience. They learn additional foreign languages at proficiency levels ranging from elementary to upper-intermediate/advanced levels. A group of participants who have not studied any additional foreign languages after L2 English is also included as a control group. LA is measured through the LLAMA test battery version 3 (Meara & Rogers, 2019). WM is measured with non-linguistic verbal (operation span) and visuospatial (rotation and symmetry span) complex WM tests. A forward digit span task as a measure of verbal PSTM is included to probe the LLAMA D subtest in the construct is supposed to represent due to reservations about its internal validity (Bokander & Bylund, 2020) and previous findings indicating its close relationship with measures of PSTM (Granena, 2013a). The predictive power of LA and WM on L3 comprehension is probed on the participants' listening and reading comprehension scores on proficiency tests in German, Spanish, French, and Italian languages. The linguistic backgrounds in L1 Turkish and L2 English are also checked with some subtests of the Woodcock-Johnson III achievement test battery in the Turkish-adapted version and Nation and Beglar's (2007) English vocabulary size test.

3.1 Participants and the context of the study

The participants are one hundred and eleven undergraduate students in an English language teaching program at an English-medium instruction university in Turkey. The entire sample includes 82 female and 29 male participants with a mean age of 21.49 years (*SD* = 1.62). An L3 learning context is obtained as the majority of them are L1 speakers of Turkish with extensive English as a foreign language learning experience through primary to higher education. In the national education curriculum implemented at primary and secondary schools, English language instruction starts in the fourth grade and includes varying hours of English language instruction through high school. At the university in the study context, the students are required to pass the university's in-house English language proficiency exam or present an equivalent of the minimum passing score on international standardized tests (550 on TOEFL PBT, 213 on TOEFL CBT, 79-93 on TOEFL IBT, and 6.5 on Academic IELTS). All the academic departments and programs conduct content teaching and assessment procedures in English. Therefore, the participants are within an academic environment with substantial use of English as a second language.

A wide range of additional foreign language courses (e.g., European, far and middle-eastern, or Slavic languages) are offered at the university. Mostly categorized as elective courses, these are open to all students from all departments. These language courses are mainly offered in three proficiency levels divided into two: codes 101 -102 (Elementary 1-2), codes 201 (Intermediate 1-2), and codes 301 (Upper-intermediate 1-2).

The foreign language classes meet three times a week for 2 hours, composing six hours of instruction per week. Most students start taking the courses from code 101 and continue up to 302 levels to achieve a functional proficiency level. The

present study included cross-sectional participants from different proficiency levels in any foreign language, coding the number of semesters they took their additional foreign language as a variable labelled as “domain experience” following Payne, Kalibatseva, and Jungers (2008). The participants who have taken a 201-level course in Spanish, German, French or Italian took listening and reading sections of proficiency tests available in these languages by certified institutions. The voluntary participants taking or having finished a 102-level course were also given the chance to take the test. The data from those who completed the tests were included in the dataset labelling the language course level for all the L3 learning participants. These four European languages are designated to measure L3 comprehension since comparable proficiency tests are available for these languages. Participants taking other L3s are also accepted to the study and categorized as L3 learners and are included in the analysis of the cognitive variables. Participants who have not taken any additional foreign languages are categorized as an L2-only group, and they were included in the analyses as the control group. Students from sophomore to senior years are accepted to participate in the study. The number of participants in each L3 domain experience and their level in the undergraduate program are presented in Table 1.

3.2 Instruments

3.2.1 Background questionnaire

A background questionnaire was developed within the context of the current design to collect data about the participants’ language history (L2 proficiency score on a standardized test, length of language instruction for L2 and L3, any experience in a

country where L2 and L3 are spoken, self-ratings of proficiency in L2 and L3), reported language use for each language daily is implemented (see Appendix A).

Table 1 Participant Numbers in Semester and L3 Proficiency Groups

Participant Profile	*N*	Per cent
Number of semesters at the time of data collection		
2 (Freshman)	1	.9
3 – 4 (Sophomore)	42	37.8
5 – 6 (Junior)	37	33.3
7 – 12 (Senior)	31	27.9
Total	111	100
Levels of L3 courses taken	*N*	Per cent
0 (L2-only)	25	22.5
1 – 2 courses (Elementary, 101-102 levels)	42	37.8
3 – 4 courses (Intermediate, 201-202 levels)	26	22.5
5 – 8 courses (Upper intermediate-advanced, 301-302 and upper levels)	19	16.2
Total	112	100
The distribution of languages as L3 courses	*N*	Per cent
Chinese	7	8.14
French	23	26.74
German	11	12.79
Italian	2	2.33
Japanese	3	3.49
Korean	9	10.47
Portuguese	9	10.47
Russian	5	5.81
Spanish	17	19.77
Total	86	

3.2.2 LLAMA Language Aptitude Test Battery

As a computerized freely available language aptitude test battery, LLAMA version 3 (Meara & Rogers, 2019) is used as the language aptitude measure (available online at https://www.lognostics.co.uk/tools/LLAMA_3/index.htm). As an updated version of LLAMA version 2 (Meara, 2005), version 3 has some improvements in item response types and test procedures. Common to both versions, auditory materials are adopted from unfamiliar languages (e.g. a dialect spoken in northern Canada or an Indian language spoken in British Columbia, etc.), aiming to reflect natural language

features while creating a novel language to deal with. Test phases on all subtests are not time-restricted. The final scores are displayed at the end of each test phase. As a difference to version 3, no feedback is provided upon response to the test items. All subtests are scored on a 0 – 20-point range.

LLAMA B is a vocabulary learning test that involves learning the names of novel objects that are presented as images on a screen. A study session of 120 seconds for 20 object items is allowed before the test session; the names of the objects are displayed when the objects are clicked on. When the test time starts, the display of object names is deactivated. In the updated version, the positions of the objects in the study phase change in the test sessions to prevent memorization through the location. A test item display button appears for the test phase and a single object name appears on each click as a written word. The test taker is expected to remember and click on the object, the name of which appears at the bottom of the page.

LLAMA D is a sound recognition test to examine whether the test taker recognizes spoken words when provided in a list mixed with unfamiliar words. The sounds in this part are computer-generated based on an Indian language. In version 3, ten target words to be recognized in the rest of the list are played in an uninterrupted order at the beginning without an explicit note to the test taker. After the first ten words, a mixed list of novel words and two repetitions of the first ten words is played. The test taker responds to every single word reporting whether that word was played earlier in the task or not by clicking on a button. The test taker receives a score for each correctly identified familiar or unfamiliar word.

LLAMA E is a sound-symbol correspondence test examining the ability to analyze the correspondence between the written and auditory representations of two-

syllable words. A study phase of 120 seconds allowed test takers to learn how the given spellings are pronounced by clicking on the written words. In the test phase, the test taker listens to the pronunciation of a two-syllable word and recognizes the correct spelling from among the spellings of all 20 test items.

LLAMA F is a grammatical inferencing test examining the ability to recognize grammatical rules in a novel language and produce new phrases applying these discovered rules. In the study phase of five minutes, the test taker examines 20 images along with phrases describing them in terms of colour, shape, number and position in a novel language following certain grammatical agreement rules. After the study phase, novel pictures showing different combinations of shapes are presented. The test taker is expected to form the phrase describing the novel picture by following the agreement rules and using the correct words. The content words and articles are provided as buttons, the test taker clicks on the buttons in the order s/he wishes to form the correct phrase.

3.2.3 Working memory span tasks

Non-linguistic verbal (Operation Span) and visuospatial (Rotation and Symmetry Span) complex WM span tasks are implemented to measure WM. Shortened versions of these span tasks by Foster et al. (2015) are downloaded from https://englelab.gatech.edu/shortenedtasks.html and administered using E-Prime 3.

The operation span (OSpan) task involves judging the correctness of arithmetic operations while remembering words that appear after every arithmetic operation. After responding to the operation stimuli, the screen is blanked for 500ms and the letter to be recalled is presented. The test includes five levels from three to

seven stimuli presented in each trial. One block includes one trial for each level, producing 25 stimuli to remember in total. The final score is calculated by summing the number of letters recalled in the correct order.

The rotation span (RotSpan) task involves deciding whether capital letters "G", "F", or "R" are normal or mirror-reversed. They are also presented rotated at various degrees (0°, 45°, 90°, 135°, 180°, 225°, 270°, or 315°) on the screen. The task is to rotate the letters mentally and decide whether they are normal or mirror-reversed clicking "yes" or "no" buttons. Half of the time the letters are normal. After responding to the rotation stimuli, the screen is blanked for 500ms and a short or long arrow radiating out from the centre of the screen pointing at 0°, 45°, 90°, 135°, 180°, 225°, 270°, or 315° directions appear on the screen for 1s followed by a blank screen for 500 ms. These rotation task – arrow stimuli sequences repeat two to five times in one block, producing 14 stimuli to recall. The final score is calculated by summing the number of arrows recalled in the correct order.

The symmetry span (SymSpan) task involves deciding whether abstract shapes are symmetrical on their vertical axis as the processing task. The storage task requires recalling the locations of red cells on a 4 x 4 grid. The levels range from two to five, producing 14 stimuli to recall in one block. The final score is calculated by summing the number of red cell positions recalled in the correct order.

3.2.4 Forward digit span task

Representing a simple recall task, this is implemented to measure PSTM. Tapping the participants' simple storage capacity, the task presents series of 3 to 9 digits. After the sequence of digits is presented, the test takers type the digits they recall in

the order they appeared on the screen. The scores on this test ranged between 3 and 9. The task was conducted using E-Prime 3.

3.2.5 L3 comprehension tests

Language proficiency tests created by certified international language and culture institutions based on the Common European Framework of Reference for Languages (CEFR) are administered to obtain a sample of L3 comprehension. To refer to substantial L3 knowledge, participants who have studied their L3 in at least pre-intermediate/intermediate took listening and reading comprehension tests in their L3. This language level is designated based on the university's School of Foreign Languages, which offers foreign language courses basically adjusted to the CEFR levels. A language course with 101 code mainly corresponds to the A1 level in the CEFR and 102 code corresponds approximately to the A2 level. The languages selected as the L3s in the current study (i.e., German, Spanish, French and Italian) are all offered by the School of Foreign Languages, and the proficiency levels across different languages are comparable. Therefore, the minimum course level to qualify to take the L3 comprehension tests is designated as 201 level, which corresponds to pre-intermediate/intermediate level (B1 level in CEFR). To tap the minimum base for L3 proficiency, language proficiency certificate tests in the B1 level in CEFR for all four languages are administered.

The scoring of the L3 comprehension tests is based on the scoring criteria as described in the tests. The participants' scores are converted to percentages to obtain a common scale across the four languages.

Table 2 Language Proficiency Tests for L3

Language test	Listening	Total items	Reading	Total items
German (ZERTIFIKAT B1 of Goethe Institute)	Part 1: Five short listening texts, 10 items Part 2: One longer text, 5 MC items Part 3: One longer text, 7 True-False items Part 4: One discussion, matching statements with speakers in 8 items	30	Part 1: One text, 6 True-False items Part 2: Two texts, 6 MC items Part 3: Matching situations with announcements, 7 items Part 4: Matching paragraphs with people's opinions, 7 items Part 5: One short text with 5 MC items	30
Spanish (DELE, Level B1)	Part 1: Five short listening texts, 5 MC items Part 2: One longer text, 6 MC items Part 3: Six short texts, 6 MC items Part 4: Six short texts, 6 matching items Part 5: One conversation, matching statements with speakers in 6 items	30	Part 1: Six short texts and personal introductions, 6 matching items Part 2: One longer text, 6 MC items Part 3: Matching paragraphs with people's backgrounds, 6 items Part 4: Insert missing sentences into a longer text, 6 items Part 5: A cloze test with 6 MC items	30
French (DELF, Level B1)	Part 1: One text, 2 short-answer items, 4 MC items Part 2: One text, 3 short-answer items, 3 MC items Part 3: One text, 3 short-answer items, 5 MC items	20	Part 1: One text, 10 True-False items Part 2: One text, 5 MC items, 6 open-ended items	21
Italian (CILS, Level B1)	Part 1: Seven short listening texts, 7 MC items Part 2: One longer dialogue, 7 MC items Part 3: One longer text, choosing statements mentioned in the text among others not mentioned, 13 items	27	Part 1: One text, 7 MC items Part 2: One text, selecting statements mentioned in the text among those not mentioned, 15 items Part 3: Putting a paragraph into order from scrambled sentences, 10 items Part 4: A cloze test with 15 items	47

3.2.6 Woodcock-Johnson III achievement test battery

This test battery includes achievement and general cognitive ability tests as measures of general intelligence, specific cognitive abilities, language, and academic achievement (Schrank et al., 2001). The subtests implemented in this study are adapted to the Turkish language (Erçetin, Babür, & Haznedar, 2014), which is used

as a control background variable for the participants' L1 Turkish linguistic abilities. The story recall (subtest 3), delayed story recall (subtest 12), and vocabulary knowledge (subtest 17) are implemented for this purpose.

The story recall taps language development and listening ability. The participants listen to short stories played on a computer. When the story ends, the participant tells the story back with as many details as s/he can remember. Points are given to chunks of information matching the actual story in meaning with exactly the same or very close vocabulary.

The delayed story recall subtest taps meaningful memory. The participants are asked to retell the stories from subtest 3 after 30 minutes-8 days. The scoring is done in the same way as subtest 3; points are given to chunks of correctly recalled information.

The vocabulary knowledge subtest taps reading vocabulary, verbal/printed language comprehension, and lexical knowledge. The subtest consists of three sections: supplying synonyms in one section and antonyms in another section for prompted words. In the third section, the participant is given groups of words that are semantically related with a blank slot to complete following the semantic relation in the preceding words. The participant completes the word groups providing the missing word.

All the subtests are conducted at the participants' pace without strict time pressure. However, they are reminded not to take more than a couple of seconds to reflect on the items, trying to keep their own pace while responding to the items. The groups formed based on L3 proficiency as indicated in Table 1 are compared for their L1 linguistic background measured with these subtests. A multivariate analysis

of variance with L3 proficiency groups as an independent variable and the three subtests as dependent variables indicated the groups do not significantly differ in any of the three subtests, F (3, 106) = .970, $p > .05$; Wilk's $\Lambda = 0.921$, providing evidence for a comparable sample in terms of L1 linguistic resources.

3.2.7 English vocabulary size test

Nation and Beglar's (2007) vocabulary size test representing fourteen thousand word families in English is implemented to represent a sample of the overall L2 English proficiency of the participants. Every 1000th-word family in a graded list of 14 frequency levels is represented with 10 multiple-choice questions. The scores of this test are used as a control variable for the participants' background L2 linguistic resource. The most common words 2000 word families are excluded since they do not promise any discriminating potential across the participants. The items starting from the 3rd frequency level are presented to the participants, leaving 120 multiple-choice items in total to be responded to. A one-way analysis of variance across the four L3 experience groups shows that there is not statistically significant overall difference in terms of L2 English vocabulary knowledge across the four groups, F (3, 106) = .183, $p > .05$.

3.3 Design and procedure

The researcher invited potential participants to contribute to the study with an overall summary and duration of experiment sessions via an email explaining the background of the study and what is included in participation. The study was also introduced in some departmental courses and volunteers were invited to participate in

their own time. Participation was not obligatory in any academic context or did not offer any financial tokens. The participants who agreed to contribute to the study were invited to individual sessions taking around 2.5-3 hours with the researcher. They were asked to fill in an online background questionnaire form before attending their session. The session started with the Woodcock-Johnson subtests 3 and 17 subtests, taking around 10-15 minutes. Afterwards, the forward digit span task was administered, taking around 5 minutes. These are followed by the WM span tasks follow with around 40 minutes to complete three blocks of symmetry, 2 blocks of rotation, and 1 block of operation span task, implementing Foster et al.'s (2015) shortened models accounting for comparable variability in general intelligence with full versions with more efficient time for participants to complete. The four subtests of the LLAMA tests follow these sections, with a fixed order of LLAMA B, D, E, and F. This is followed by the delayed story recall test of Woodcock-Johnson subtest 12 and the English vocabulary size test is administered as the last instrument. The L3 learners who could take the L3 comprehension tests complete them after all the preceding tests common to all participants. The participants are offered a ten-minute break if they need it before moving on to the vocabulary size test. The researcher provides all the instructions in the participants' L1 Turkish. The cognitive tasks are administered in a fixed order (i.e., digit span, WM tasks and the LLAMA). In cases of sessions clashing two or more participants for 15-30 minutes in the computer laboratory, participants take the vocabulary size before all the other tests to avoid distraction from other participants during the cognitive tasks. Approval of the ethics committee was also obtained (Appendix B).

3.4 Data analysis

The test variables were initially checked for missing data. Three participants' data were removed due to missing sessions in either all the LA or WM measures. The remaining 108 participants' data were then checked for univariate normality statistically and visually with histograms and box and whisker plots. Some zero points were observed in LLAMA B, D, and E tests. These observations were checked for their impact on the overall distributions of the variables. Two cases with zero points standing above - 3 standard deviations (SD) below the mean in LLAMA D were removed from the dataset. The zero scores observed in LLAMA B and E were within the -/+ 3 SD range, therefore were kept in the dataset. The distributions of the LLAMA subtests did not yield any extreme or outlier cases after these processes. The skewness and kurtosis values were checked against the critical -/+ 1.96 value following Field (2013), establishing evidence for univariate normality. L3 listening and reading comprehension scores were converted to percentages from their original scales. The univariate normality was confirmed for these two variables with skewness and kurtosis values, Kolmogorov-Smirnov tests of normality, and histograms.

OSpan scores displayed highly negative skewness (skewness = -1.18, std. error = .234). A log 10 transformation on the reflected scores of OSpan was implemented, which effectively remedied the univariate non-normality in this measure. SymSpan scores displayed a leptokurtic distribution (kurtosis = 1.138, std. error = .463 with slightly negative skewness. This deviation from normality was remedied with a square root transformation on the reflected scores of SymSpan. RotSpan scores included data from one block of the task since eleven participants

could complete only on block o the task due to technical issues. To include their data in the analysis, the first block of Rotspan was included for the entire dataset.

The remaining data were checked for multivariate outliers using the Mahalanobis distance criterion of $p < .001$ across the complex WM, forward digit span tasks, and the LLAMA subtests, detecting no multivariate outliers in the dataset.

The first research question investigated the relationships among the measures of LA and WM and probed their construct-level descriptions. Bivariate Pearson product-moment coefficients among the four subtests of the LLAMA, complex WM tasks and the forward digit span task were examined. Then, an exploratory factor analysis (EFA) with a principal component technique was conducted. Bartlett's test of sphericity confirmed that the correlation matrix was not random, χ^2 (28) = 121.02, $p < .001$. The obtained Kaiser-Meyer-Olkin measure of sampling adequacy was .649, which was above the acceptable minimum value of .50 (Field, 2013). The derived factors were saved as variables using the regression method.

The second research question investigated the influence of L3 learning on the factors derived from LA and WM measures represented as the factors in the EFA. Univariate analyses of variance using one-way ANOVAs on the three factors derived from the EFA were conducted to examine between-subjects effects. The error variance in the dependent variables across the groups of L3 learning was confirmed with non-significant results of Levene's tests. The homogeneity assumption was confirmed with the normality of the studentized residuals.

The third research question investigated the predictive roles of the EFA factors on L3 listening and reading comprehension. Hierarchical linear regression analyses were conducted controlling for the level of L3 learning experience to

examine the additional contributions of the factors in explaining variance in L3 comprehension scores. The multicollinearity assumption was checked with VIF statistics and tolerance. Homoscedasticity was checked with the distributions of the residuals of the predicted values.

CHAPTER 4

RESULTS

This chapter presents the results of the statistical analyses addressing the research questions. The results are presented following the order of the research questions. First, the relationships between the LA and WM measures are presented. Then, the analyses differentiated by the L3 domain experience are presented in relation to the LA and WM factors. Finally, the predictive roles of the LA and WM factors in explaining L3 comprehension scores are presented.

4.1 LA and WM relationships

The descriptive statistics of the variables for the entire sample are presented in Table 3. A pairwise exclusion of the missing observations for the participants who could not complete some of the tasks was implemented in the descriptive statistics to provide an overall view of the data. Therefore, the number of participants for each variable is provided in the table. The highest mean score across the subtests LLAMA was observed in LLAMA F, while LLAMA B displayed the lowest mean scores. The descriptive statistics divided by the groups of L3 learning experience as presented in Appendix C.

Table 3 Descriptive Statistics for the Entire Sample

	N	*M*	*SD*	*Min*	*Max*	*Skew.*	*Kurt.*
LLAMA B (Possible max. = 20)	108	9.26	4.39	0	19	.34	-.58
LLAMA D (Possible max. = 20)	106	11.34	2.78	3	18	-.36	.74
LLAMA E (Possible max. = 20)	108	11.28	5.18	0	20	-.44	-.83
LLAMA F (Possible max. = 20)	108	12.09	3.6	4	19	.03	-.70
OSpan (Possible max. = 25)	107	20.64	4.09	7	25	-1.18	1.17
RotSpan (Possible max. = 14)	105	8.96	3.03	0	14	-.51	-.21
SymSpan (Possible max. = 42)	107	28.82	7.82	0	42	-.78	1.14
Digit span (Possible max. = 9)	108	7.13	.99	4	9	-.34	.41
L3 reading comp. percentage	34	58.81	15.28	30	84	-.02	-1.17
L3 listening comp. percentage	34	55.86	15.27	26.67	90	.34	-.01

The Pearson product-moment coefficients for the correlational analysis among the LA and WM measures are presented in the correlation matrix in Table 4. The correlation matrix indicated that LLAMA D (i.e., phonetic coding) and digit span (i.e., PSTM) were significantly correlated with a weak magnitude of .21. Additionally, SymSpan (i.e., complex visuospatial WM) had significant but weak correlations with LLAMA E (i.e., sound-symbol correspondence) and LLAMA F (i.e., grammatical inferencing). LLAMA E was also significantly but weakly correlated with OSpan (i.e., complex verbal WM). The intercorrelations among the subtests of LLAMA were also weak to moderate and significant. LLAMA B (i.e., vocabulary learning) had weak significant correlations with LLAMA D and E, and a moderate significant correlation with LLAMA F. LLAMA E also had a significant weak correlation with LLAMA D and a slightly stronger correlation with LLAMA F. The only non-significant correlation was observed between LLAMA D and F.

Table 4 Correlation Matrix for The LA and WM Measures

	1	2	3	4	5	6	7	8
1. LLAMA B	1							
2. LLAMA D	.206*	1						
3. LLAMA E	.269**	.205*	1					
4. LLAMA F	.454**	.184	.348**	1				
5. OSpan	-.055	.015	.256**	.153	1			
6. RotSpan	.144	.059	.129	.128	.225*	1		
7. SymSpan	.158	.184	.239*	.323**	.266**	.487**	1	
8. Digit span	.101	.210*	.075	.148	.260**	.106	.247*	1

Note: *$p < .05$, **$p < .01$

The LA and WM measures were entered into EFA. Bartlett's test of sphericity confirmed that the correlation matrix was not random, χ^2 (28) = 121.02, $p < .001$. The obtained Kaiser-Meyer-Olkin (KMO) measure of sampling adequacy was .649, which was above the acceptable minimum value of .50 (Field, 2013). The results of the EFA with varimax rotation produced three factors with eigenvalues above 1. The three factors explained 30.78, 15.73, and 12.96 % of the variance respectively with a total variance of 59.57%. The communalities ranged between .411 and .754. The varimax-rotated factor loading scores are presented in Table 5. The factor structure was probed with a direct oblimin rotation method, displaying the same factor structure with the same variables loading with the same factors as the varimax-rotated solution. The factor loadings of the LA and WM tests in indicated that the LLAMA D sound recognition test had quite close factor loadings above .4 on both factor 1 with the other LLAMA subtests and factor 3 with OSpan and digit span tasks.

Table 5 Rotated Factor Loadings

	Factor 1	Factor 2	Factor 3
LLAMA B	.816	.077	-.157
LLAMA D	.467	-.197	.440
LLAMA E	.557	.166	.271
LLAMA F	.753	.183	.090
OSpan	.017	-.339	-.674
Rotspan	.064	.866	.015
SymSpan	-.252	-.743	-.244
Digit span	.082	.043	.765

To probe the consistency of the model without the LLAMA D, the same factor analysis was conducted excluding the LLAMA D. This model also met the assumptions with a KMO value of .609 and a significant Bartlett's test of sphericity, χ^2 (28) = 107.91, $p < .001$. The model displayed the same factor model with the same variables loading onto the same factors. The three factors explained 32.82, 17.67, and 14.57 % of variance respectively with a total variance of 65.06 %, slightly improving the previous model with the LLAMA D in terms of explaining variance. However, the factor scores saved from this second model displayed similar results in the following analyses as the model including the LLAMA D. Studies on factor analysis in social sciences offer some checkpoints to decide on cross-loading items such as high factor loading of the other variables and presence of at least two variables representing the factors other than the cross-loading factor (e.g., Costello & Osborn, 2005; Matsunaga, 2010). Based on the obtained evidence meeting these check points, the factor model including the LLAMA D scores was used and reported for the following analyses to represent a more complete picture with the available data.

Spearman's rho coefficients were calculated to examine bivariate correlations among the factor scores as seen in Table 6. All the correlations were at very small magnitudes and non-significant.

Table 6 Spearman Correlations among the Factor Scores

	1	2	3
1. Factor 1 (LLAMA)	1		
2. Factor 2 (visuospatial WM)	.018 (.854)	1	
3. Factor 3 (verbal/phono. memory)	.017 (.861)	.032 (.752)	1

Note. Two-tailed *p* values are provided in parentheses

Despite not reaching statistical significance, the magnitudes of the correlation coefficients indicated a slightly higher correlation between the visuospatial and verbal/phonological memory factors. On the other hand, factor 1 with the LLAMA subtests displayed a clearer dissociation with comparably small correlations with both visuospatial WM and verbal/phonological memory factors.

4.2 Relationships of L3 experience with LA and WM factors

The second research question examined the relationships of the L3 domain experience, defined as the number of semesters the participants attended their L3 courses, with the LA and WM measures. Considering the correlational dissociation among the factor scores, separate univariate one-way ANOVAs on the three factors were conducted to examine the relationships of the L3 experience across its four levels (i.e., L2-only, beginner, intermediate, and upper-int./advanced) on the factor scores. The assumption of homogeneity across the levels of the L3 experience was

confirmed with non-significant results of Levene's tests for the three dependent variables. The assumption of homoscedasticity was checked with the normality of the standardized residuals, confirmed for all three dependent variables. Table 7 indicates that the L3 experience did not have any significant effects on any of the factor scores representing the LLAMA, visuospatial WM, and verbal/phonological memory measures.

Table 7 Tests of Between-Subjects Effects for Factor Variables

Source	Dependent variable	*SS*	*df*	*MS*	*F*	Partial η^2	*p*	Observed power
L3 experience	Factor 1 (LLAMA)	5.986	3	1.995	2.058	.059	.11	.513
Error		96.014	99	.970				
L3 experience	Factor 2 (visuospatial WM)	3.745	3	1.248	1.258	.037	.29	.327
Error		98.255	99	.992				
L3 experience	Factor 3 (verbal/phono. memory)	3.831	3	1.277	1.288	.038	.28	.335
Error		98.169	99	.992				

To probe the robustness of the non-significant results of the overall *F* tests in the ANOVA models, orthogonal contrasts for the effect of the L3 experience in the dependent variables were also checked as the final step of the analysis. Planned contrasts confirmed the non-significant results of the omnibus *F* tests. The L2-only group did not differ from the three L3 learning groups in the LLAMA factor 1, t (99) = -1.056, p = .294. The beginner L3 group did not differ from the remaining two groups, t (99) = -1.247, p = .215. The intermediate and upper-int./advanced groups did not significantly differ, either, t (99) = -1.807, p = .074.

In the visuospatial WM factor, the L2-only group did not significantly differ from the L3 learning groups, t (99) = 1.028, p =.306. The beginner L3 group did not differ from the other two L3 groups, t (99) = .176, p = .860. The intermediate and the upper-int./advanced groups did not differ, t (99) = 1.587, p = .116.

The verbal/phonological memory factor displayed the same contrasts, indicating no significant difference for the L2-only group from the L3 learning groups, t (99) = -.432, p =.667. The beginner L3 group did not differ from the other two L3 groups, t (99) = .340, p =.735. The intermediate and the upper-int./advanced groups did not differ, t (99) = -1.877, p = .064.

4.3 LA and WM factors in predicting L3 comprehension

The third research question investigated the predictive roles of the LA and WM factor scores on the L3 listening and reading comprehension scores. The listening and reading comprehension scores were analyzed as separate dependent variables in linear hierarchical regressions controlling for the L3 learning experience at the first step. The multicollinearity assumption was checked with VIF and tolerance statistics. The observed variance inflation factor (VIF) values were below the critical value of 10 and tolerance statistics were all above the critical value of 0.1 as suggested in Field (2013). The homoscedasticity assumption was checked with the normality of the standardized residuals and confirmed for both regression analyses. Table 8 presents the Pearson correlations between the L3 listening comprehension as the dependent variable and the L3 group, LLAMA, visuospatial WM, and verbal memory factors as the independent variables.

Table 8 Correlation Matrix for the L3 Listening Comprehension and the Independent Variables

	1	2	3	4	5
1. L3 listening comprehension	1				
2. L3 experience	.545***	1			
3. Factor 1 (LLAMA)	.282	-.039	1		
4. Factor 2 (visuospatial WM)	-.045	.093	-.130	1	
5. Factor 3 (verbal/phono. memory)	.084	-.156	-.192	-.012	1

Note: ***$p < .001$

Table 9 present the results of the analyses with L3 listening comprehension as the outcome variable. In the first step, the L3 experience explains 30% of the variance in L3 listening comprehension. The inclusion of the factor scores significantly contributed an additional 15% of variance to the model, explaining 45% of variance overall in L3 listening comprehension. When the coefficients of the variables were examined, only factor 1 representing the LLAMA B, E, and F (and LLAMA D partially) was the significant predictor in the additional contribution beyond the L3 experience.

Table 9 Results for the Hierarchical Regression on L3 Listening Comprehension

		Unstandardized coefficients		Standardized coefficients				
Steps	Predictors	*B*	*SE*	*β*	*t*	*F*	R^2	ΔR^2
1	Constant	30.58	7.23			13.52**	.3	
	L3 experience	11.62	3.16	.55	3.68**			
2	Constant	27.9	6.93			5.89**	.45	.15
	L3 experience	12.83	2.99	.60	4.28**			
	Factor 1 (LLAMA)	5.74	2.36	.35	2.43*			
	Factor 2 (visuospatial WM)	-.81	2.12	-.05	-.38			
	Factor 3 (verbal/phono. memory)	3.36	1.96	.24	1.71			

Note: *$p < .05$, **$p < .001$

Similarly, Table 10 presents the correlation matrix between the predictor variables and L3 reading comprehension. All the bivariate correlations were non-significant.

Table 10 Correlation Matrix for the L3 Reading Comprehension and the Independent Variables

	1	2	3	4	5
1. L3 reading comprehension	1				
2. L3 experience	.282	1			
3. Factor 1 (LLAMA)	.178	-.039	1		
4. Factor 2 (visuospatial WM)	-.073	.093	-.130	1	
5. Factor 3 (verbal/phono. memory)	.048	-.156	-.192	-.012	1

The regression analysis with the L3 reading comprehension as the outcome variable did not display any significant models in either step. The L3 experience as the control variable did not significantly explain any variance in L3 reading comprehension as shown in Table 11.

Table 11 Results for the Hierarchical Regression on L3 Reading Comprehension

		Unstandardized coefficients		Standardized coefficients				
Steps	Predictors	*B*	*SE*	*β*	*t*	*F*	R^2	ΔR^2
1	Constant	45.74	8.28			2.76	.08	
	L3 experience	6.01	3.62	.28	1.66			
2	Constant	43.88	8.67			1.16	.14	.06
	L3 experience	6.78	3.75	.32	1.81			
	Factor 1 (LLAMA)	3.44	2.96	.21	1.16			
	Factor 2 (visuospatial WM)	-1.12	2.66	-.07	-.42			
	Factor 3 (verbal/phono. memory)	1.88	2.46	.14	.77			

4.4 Summary of the results

The analyses addressing the first research question about the relationships among the LA and WM subcomponents revealed partial and small to moderate relationships. A PCA revealed a three-factor solution, labelled as LLAMA (LLAMA B, E, and F), visuospatial WM (RotSpan and SymSpan), and verbal memory (OSpan, LLAMA D, and digit span). The distinction between the visuospatial WM and LLAMA encompassing rote learning, grammatical inferencing, and sound-symbol correspondence was clear while the intersection point appeared to cover phonological and verbal aspects of PSTM, verbal WM and sound recognition subtest of LLAMA. These findings confirm Hypothesis 1 in that there is evidence for both distinctions and interactions between the LA and WM measures. The distinction between the verbal and visuospatial WM measures also confirmed Hypothesis 2 which predicted a separation between the verbal and visuospatial WM.

The second research question addressed the effect of L3 proficiency on the three factor scores derived from the PCA. The univariate analysis revealed non-significant results, refuting Hypothesis 3 predicting some significant effect of L3 proficiency on the LA and WM factors.

The third research question examined the predictive roles of the LA and WM factor scores in explaining L3 listening and reading comprehension scores. The results partly confirmed Hypothesis 4 in only the L3 listening comprehension scores with some predictive role of the LLAMA factor preceded by L3 proficiency. However, the regression analysis did not produce any significant models predicting L3 reading comprehension.

CHAPTER 5

DISCUSSION AND CONCLUSIONS

5.1 Findings on the LA and WM relationship

The factors derived from the PCA were labelled as LLAMA (LLAMA B, E, F, and partially LLAMA D; representing rote memory, sound-symbol correspondence, grammatical inferencing, and sound recognition respectively), visuospatial WM (RotSpan and SymSpan), and verbal memory (digit span, OSpan, and partially LLAMA D). When the factor loading scores of the variables were examined, the strongest scores on each factor were LLAMA F (factor loading score .753), digit span (factor loading score .765), and RotSpan (factor loading score .866). Assuming the latent factors are most strongly represented by these variables, the distinction among these variables appears to lie under the language-specific grammatical inferencing, more domain-general visuospatial WM, and the verbal/phonological storage capacity. The intersection point among the subtests of LA and WM involved verbal/phonological abilities. The factor loading scores of the LLAMA D sound recognition subtest displayed very close scores to both the LLAMA factor (factor loading score.467) and verbal memory (factor loading score .440), making its position more ambiguous among the set of abilities under the LA and WM.

The composition of the LLAMA factor included the more traditional LA components, namely rote memory (LLAMA B) and grammatical inferencing (LLAMA F) as observed in previous research (e.g., Granena, 2013; Xiang et al., 2012; Yalçın et al., 2016). The subtests of the MLAT representing these two abilities

were reported as the strongest variable to load onto the LA latent factor (Winke, 2013).

The factor loading of RotSpan and SymSpan as complex visuospatial WM were in line with the predictions of the study, which assumed to represent a non-linguistic domain-general cognitive ability to probe the role of domain-general learning mechanisms in L3 learning following the arguments in L2 learning (e.g., DeKeyser 2000; Skehan, 2019). Therefore, the factor representation by these two WM measures appears consistent with the hypothesis which predicted that visuospatial WM would constitute a distinct latent variable.

The verbal memory factor constitutes a more complex structure regarding the types of cognitive abilities underlying the tasks. The considerably strong loading of OSpan could be evaluated based on Foster et al.'s (2015) caution on its reduced predictive role in explaining general intelligence on its own without the contributions from rotation and symmetry span tasks (p.234). In other words, OSpan in its full block version was not observed to be a strong enough predictor of domain-general intelligence, raising questions about its validity as a measure of domain-general WM ability. The findings in this study could also be providing support for this caution on the OSpan as a measure of WM considering its loading onto a common factor with PSTM. Digit span representing PSTM, the storage component of WM, would be expected to represent a different structure from complex verbal WM represented with OSpan. Although slightly above the cut-off point of .40, the loading of LLAMA D onto this factor also appeared to share the verbal storage aspect with this factor. Considering the verbal storage aspect shared in all tree tasks, this factor could be representing verbal storage as the latent factor with the strongest loading from the digit span task.

At this point, it might be important to reflect on the research and arguments on the essence of the LLAMA D task, regarding what it might represent as a construct. Granena's (2013a) findings indicating the LLAMA D sound recognition subtest loading onto a separate factor from the other three LLAMA subtests and with the serial reaction time task used as a measure of implicit learning, was interpreted as LLAMA D representing a separate factor which could be more implicit in nature with less involvement of analytical attentional processes since it lacked a separate study phase. The other three subtests all involved a preceding study of the input material before the test phase, which was argued to allow an opportunity to reflect consciously on the material and activate more explicit processing above the consciousness level. Following studies also largely confirmed this distinction of LLAMA D from the other subtests, with LLAMA B rote memory and LLAMA F grammatical inferencing subtests as more consistent subtests to load together (e.g., Artieda & Munoz, 2016; Granena, 2019; Yalçın et al., 2016).

However, studies probing the construct structure of LLAMA D pointed to some areas of caution in interpreting LLAMA D as a measure of implicit processes. For instance, Suzuki (2021) implemented some revisions to the instructions of LLAMA D and measured reaction times and confidence ratings of the participants about their responses to each item along with accuracy scores. The instructions were revised not to direct the participants' attention to the similarity or difference between the first ten words in the exposure phase of the LLAMA version 2. Instead, the participants were directed to focus on the sound volume while the exposure phase was played. Suzuki compared the confidence ratings with accuracy and reaction time and found that both accuracy and reaction time scores were positively related to confidence rating. This was interpreted as participants' implementing conscious

processing while deciding on their responses and that their consciousness had an effect on their accuracy and reaction time, challenging the potential of LLAMA D as a measure of implicit aptitude aspect. Suzuki emphasized the potential of LLAMA D as a measure of proceduralisation through reaction time, as proceduralisation did not necessarily require unconscious processing. Therefore, he argued that LLAMA D could be used as "a measure of a memory-based, procedural system", particularly for oral skills (Suzuki, 2021, p.673).

Iizuka and DeKeyser (2023) also reported some challenging findings based on instruction types regarding the position of LLAMA D as an implicit aptitude measure in relation to other measures of implicit learning. Three conditions of instruction were created, i.e., memorisation, just-listen, and sound-check conditions. The memorisation condition followed the original instructions as Meara and colleagues exemplified in the test manuals, activating focal attention, intention, and awareness, inducing a conscious approach to the test. In the just-listen condition following Granena's (2013a) procedure, the participants were not told about the following test phase. They were instructed to respond based on their familiarity in the test phase. In the sound-check condition following Saito's (2017) procedure, the participants were just told that the exposure phase was just a sound check to see if they could hear any sounds while the exposure phase was played. They were then instructed to make judgements based on familiarity in the test phase. The LLAMA D scores under these three instruction conditions were then regressed on other measures of implicit and explicit memory or executive functions, i.e., serial reaction time, digit span, available long-term memory, paired associates, Stroop task, and sound discrimination task. The results showed that the LLAMA D scores under the sound-check condition were not predicted by any of the other implicit, explicit or executive

function variables. In the memorisation condition, only the paired associates test measuring rote memorisation was a significant predictor with a negative direction, while in the just-listen condition, serial reaction time scores significantly predicted LLAMA D scores. These findings indicated the critical impact of the instruction type while using the LLAMA D with possibilities of reflecting different points of the explicit-implicit continuum of processing. DeKeyser (2019) also highlighted that LLAMA D could be presenting a sketch of language processing from the "implicit end of the spectrum" rather than a dichotomous distinction between explicit and implicit processing (p.320).

Returning to the context of the current study where the new version 3 of the LLAMA tests was used, the slight adaptation in the exposure and test phases could have created some differentiated impact in this regard. Version 3 of LLAMA D presented all the familiar and unfamiliar words in an uninterrupted string, without signalling the initial exposure phase. Therefore, the participants in this study were instructed to listen to the words and to respond to each word, deciding whether they had heard that word in the string of words they hear in the task. They were particularly directed to confine their comparison to the particular LLAMA D test rather than a wider search in all languages they were familiar with. This procedure could place the instruction in this study somewhere between Granena's "just-listen" and Meara and colleagues' original "memorisation" instruction as the focus had not been completely distracted from the test session as in the "sound-check" instruction. Considering these factors, the results indicated that LLAMA D stood in between the "explicit" subtests of LLAMA and the memory measures using phonological/verbal input material, namely, digit span and OSpan. Therefore, it could be argued that LLAMA D could represent the area where the domain-general verbal memory and

the domain-specific LA (processing particular language material) could potentially overlap. Factor 3 labelled as verbal/phonological memory included the tasks with verbal input material rather than differentiating the type of processing they involved. The processing of verbal material in these tasks was mostly conscious following the theoretical conceptualisations of WM especially from unitary perspectives with working memory representing the activated attention unit (e.g., Cowan, 2005; Engle, 2002). Although LLAMA D did not have a very strong loading score with this factor slightly going beyond the cut-off point of .40, it is the only LLAMA subtest to load apart between the LA and memory factors. Therefore, the processing of the verbal material in this task could also entail processing at the conscious level as in the digit span and OSpan tasks.

As the strongest variable loading onto this verbal/phonological memory factor with a loading score of .765, digit span representing PSTM was previously observed to be a significant predictor of high proficiency learners' L2 listening and reading comprehension scores in the Hi-LAB study (Linck et al. 2013), measuring it with letter span and non-word span tasks. The other significant predictors were serial reaction time representing implicit learning and paired associates representing associative memory. OSpan followed the digit span task with a factor loading score of -.674 with a negative direction due to the reflection applied to the scores in the data transformation. OSpan involved a distracting processing task along with storage capacity involving both executive function and storage functions of WM. Unsworth and Engle (2005) investigated the role of OSpan in incidental (not intentional, out of the focal attention) and intentional (more cognitive control within focal attention) learning conditions. They observed effects of high and low WM only in the intentional but not in the incidental learning condition, reporting evidence for the

more explicit and attentional structure of the OSpan task. Therefore, in addition to the input material, the processing underlying the three tasks in the verbal/phonological factor could be sharing some common ground in terms of attentional level. The bivariate correlations between the tasks also indicated that only digit span significantly correlated with both OSpan (r = .260, p < .001) and LLAMA D (r = .210, p < .05 while OSpan and LLAMA D did not significantly correlate. This view of the three tasks could be pointing to a mediating role of PSTM between latent variables underlying OSpan and LLAMA D tasks, deserving attention with a specific research design. Understanding this intersection in detail could provide empirical evidence regarding the propositions of the recent theoretical approaches (e.g., Wen & Skehan, 2021) to the roles of phonological/executive cognitive abilities in the different stages of L2 learning.

5.2 Relationships of L3 experience with LA and WM factors

The findings of this study indicated that the L3 experience representing L3 proficiency-based representation of the number of semesters the participants continued their additional foreign languages, did not have any significant relationships with the LLAMA, visuospatial WM or verbal/phonological memory. These non-significant results counteract previous studies reporting relationships between L3 proficiency and measures of LA (e.g., Ma et al., 2018), visuospatial or verbal WM (Bouffier et al., 2020; Cockcroft 2022; Espi-Sanchis & Cockcroft, 2022). At this point, it is worth noting the differences in operationalization and context of multilingualism while comparing the results of the studies. Espi-Sanchis and Cockcroft's (2022) study considered the balance of proficiency across reading, writing, and speaking in the local languages of the participants who acquired them in

their home or cultural environment while the participants in the current study learned their L2 and L3 in instructed environments, with L3 experience taking place through more limited exposure to the target languages. Likewise, Bouffier et al. (2020) study also reported data from the Luxembourg context where the positions of English and French as official languages could provide different acquisition contexts and learning experiences than instructed environments. Returning to the mixed findings in the L2 research reporting various findings on L2 proficiency and LA and WM abilities (e.g., Artieda & Munoz, 2016; Hummel, 2009; Sáfár & Kormos, 2008), the multilingual environments with L3/n experience could be presenting a similar picture. There is a need for more research from different contexts with comparable instruments to probe this interpretation. Impeding factors to define and depict the multilingualism concept might need some criterion control variables such as length or amount of exposure, and the social or economic statues of the language. The sample in this study was assumed to be homogeneous in these aspects as well as L1 and L2 linguistic knowledge backgrounds. A sample with more variability in these aspects could exhibit more sensitive effects of L3 proficiency on cognitive resources such as verbal and visual WM or LA.

5.3 Predictive roles of LA and WM factors on L3 comprehension

The scores of L3 comprehension tests displayed different results concerning the three factor scores revealed in this study. For L3 listening comprehension, the LLAMA factor contributed an additional 15% of variance after L3 proficiency, explaining 45% of the variance in total, while none of the predictive models could significantly predict L3 reading comprehension. Li's (2016) comprehensive meta-analysis of LA research reported small to moderate relationships with L2 listening and overall

aptitude (r = .30), phonetic coding (r = .12), rote memory (r = .21), and spelling clues (r = .29). For L2 reading, the study reported slightly higher correlations of overall aptitude (r = .39), phonetic coding ability (r = .28), language analytic ability (r = .35), rote memory (r = .20), number learning (r = .29), and spelling clues (r = .44). While these data point to some sort of linearity between LA and L2 reading and listening measures across different studies, the variability in the types of tasks and operationalization of the measures still present some complexity in comparing these results in different studies. he validation study on the Hi-LAB test battery (Linck et al., 2013) included L2 listening and reading comprehension as L2 outcome variables and reported paired associates (available long-term memory), serial reaction time (implicit learning) and letter span (PSTM) as significant predictors of L2 listening and reading comprehension in highly advanced adult L2 users.

Studies with adolescent L2 learners also indicated moderate relationships of L2 listening comprehension to LA measures (e.g., language analytic ability, inductive language learning, rote learning) along with affective perception measures such as motivation and anxiety (e.g., Sparks et al., 2011) or metacognitive awareness of listening strategies also involving perceptions of listening anxiety and difficulty (e.g., Sok & Shin, 2022).

For the role of WM in relation to listening comprehension, Was and Woltz (2007) proposed models to explain the relationship between attention-driven WM and listening comprehension, highlighting the mediating role of available long-term memory and content knowledge. Although the model from the Hi-LAB battery included associative long-term memory as a memory component and PSTM in the prediction model for L2 listening, the factors involving PSTM or memory functions were not significant predictors of L3 listening comprehension in this study.

The prediction models with the LA and WM factors were not significant in explaining L3 reading comprehension. L3 experience as the control variable entered at the first step in the hierarchical regression was not a significant predictor, either. This dissociation was also observable in the non-significant correlations between the L3 reading comprehension and the predictor variables. This dissociation was further confirmed in the regression analysis. The lack of the L3 experience effect could be traced back to the descriptive statistics divided by the L3 proficiency groups. The elementary L3 group had descriptively higher L3 reading scores than the intermediate L3 group, pointing to a clue about the non-linearity between L3 proficiency and L3 reading comprehension. At this point, it might be important to highlight the participant profile with an extensive L2 instruction background before the L3 learning experience. Therefore, their approach to L3 reading could employ more varied linguistic and cognitive resources than an L2 context, which could neutralize the L3 proficiency effects across the L3 proficiency groups. Still, it would be expected that L3 proficiency could be associated with L3-specific baseline resources such as vocabulary and grammar knowledge which are particularly critical for bottom-up processing in complex skills such as reading before the employment of strategies and different resources. However, the L3 exposure and classroom experience might not have an impact large enough to create observable cross-sectional differences over the number of semesters that designated the L3 groups in the study. The elementary L3 group had been studying their L3 for 1-2 semesters while the intermediate group had 3-4 semesters of classroom experience, which did not seem to be associated with large differences in L3 reading comprehension.

In the L2 reading research, the role attributed to WM displayed interactions with topic familiarity and content knowledge in L2 reading comprehension in

advanced (e.g., Alptekin & Erçetin, 2011) or beginning L2 proficiency (e.g., Leeser, 2007). In the current study, the factor solution derived from the LA and WM subtests did not significantly relate to L3 reading comprehension as complex factor predictors. A more in-depth design to asses reading comprehension with more control over the reading task and other controls such as topic familiarity, content knowledge or reading text features could yield more differentiated results than a holistic approach.

The relationships of LA and L2 reading comprehension were investigated in studies involving reading as a measure of general L2 proficiency along with other L2 outcome measures. For instance, Sparks et al. (2011) reported factors involving language analytic ability and phonemic coding contributed to explaining variance in L2 reading comprehension after the self-perceptions factor involving motivation and anxiety in high school students in the US learning L2 for two years at the time of testing. Hummel (2009) reported significant correlations between L2 reading and overall aptitude as well as spelling clues and words in sentences sections of the MLAT for young adults studying English teaching French-speaking participants. Phonological memory, on the other hand, did not have any significant correlations with L2 reading but had significant moderate to low correlations with L2 vocabulary and grammar measures.

To summarize, the results of the current study indicated a clear distinction at the construct level between LA abilities (i.e., grammatical inferencing, rote memory, and sound-symbol correspondence) measured through the LLAMA version 3 and visuospatial WM measured through symmetry and rotation span tasks, challenging Miyake and Friedman's (1998) "LA as a function of WM" argument. On the other hand, the overlapping abilities categorized under LA and WM abilities were

verbal/phonological memory abilities, namely PSTM and verbal OSpan tasks partly shared by the LLAMA D sound recognition task, constituting a separate factor together. The relationships of L3 experience with these three factors were not statistically significant at any L3 proficiency levels. The predictive roles of the three factors for the L3 outcome measures displayed different pictures for the listening and reading comprehension scores. LLAMA factor was a significant contributing predictor beyond L3 experience in explaining L3 listening comprehension. However, L3 experience and the three cognitive factors were not significant predictors of L3 reading comprehension.

5.4 Conclusions

This study examined the interrelations between verbal and visuospatial components of WM and language-related abilities as represented by the subtests of the LLAMA in the L3 learning context. The results confirmed the structural distinction between the most typical subcomponents of grammatical inferencing, rote memory as well as sound-symbol correspondence and the verbal and nonverbal WM components involving processing and storage. However, the crossing point between the two constructs appeared to lie under the verbal/phonological aspects highlighting the storage aspect of the data from digit span, OSpan, and LLAMA D.

The relationships of L3 experience with the factor scores did not yield any significant results as opposed to previous findings showing positive relationships with L3 proficiency or higher intensity in the improvement of LA and WM in learning multiple languages conditions. At this point, it becomes important to clarify the participants' level of multiple language proficiencies and learning contexts to

compare and contrast findings from different studies. The current study aimed to describe L3 proficiency as indexed by the length of L3 instruction in addition to some samples of L3 comprehension from participants with a working level of L3 proficiency.

The predictive roles of the three LA and WM-based factors produced different pictures for listening and reading comprehension. The LLAMA factor was observed to significantly contribute an additional proportion of variance explained in L3 listening comprehension. Reflecting the grammatical inferencing, rote memory and sound-symbol correspondence abilities more reliably, this factor indicated that higher scores in these abilities are positively related to L3 listening comprehension. However, reading comprehension was dissociated from L3 proficiency as well as the three factors derived from LA and WM measures. This lack of correlation could be related to the measurement of reading comprehension considering the mediating rather than direct roles reported for particularly WM between reading comprehension and other predictors such as topic familiarity or content knowledge.

5.5 Implications

The conceptual and methodological implications of the findings of this study could be mostly related to the measures of LA and WM and the underlying structures they are assumed to represent. While the findings confirmed the dissociation between the more traditional LA and WM abilities, the case of LLAMA D as a sound recognition task remains in an ambiguous position regarding the latent factor it could represent. The cross-loading observed in this study implies that it could share aspects with both traditional LA abilities and the phonological/verbal storage aspects of WM. This

ambiguity could also stem from its relatively low psychometric properties as reported by Bokander and Bylund (2020) about its item discrimination indices. Although the updated version 3 introduced some corrections about the presentation of the target items, the phonological material remained the same, with no reliability or validity studies reported yet. The construct validity of OSpan as a measure of domain-general WM measure could also be an area to probe in future research. As cautioned by Foster et al. (2015), implementing OSpan as the only measure of WM could be problematic in representing the latent variable accurately, challenging the wide use of OSpan as a measure of complex WM measure.

The null results regarding the L3 experience relationships can emphasize the importance of the operationalization of multilingualism in LA and WM interaction studies. This study implemented a between-groups factor of L3 proficiency involving a cross-sectional design. However, the L3 proficiency effect as a within-group variable for multilingual individuals defining proficiencies across their languages would create a different proficiency measure as well. The observation of LA and WM over a time course in T. Huang et al. (2022) differentiated in intensity by the number of foreign languages studies presented a longitudinal perspective. In that sense, observing the LA and WM for a specified point of time could provide more nuanced findings as opposed to one-time cross-sectional designs.

Finally, one pedagogical implication could be about the development of receptive skills in multiple language learning contexts. The significant positive association of the LLAMA factor with L3 listening comprehension offers a potential set of abilities to facilitate/support the development of the listening skill while reading skill could be more dissociated from these abilities and require the

exploration of additional factors to support the development of L3 reading comprehension.

5.6 Limitations and future research

It should be acknowledged that this study is not without limitations. One point is about sampling, which mainly involved convenience sampling for the study. The participant profile could be assumed to be rather homogeneous regarding linguistic backgrounds, L2 and L3 learning contexts and experiences, socioeconomic status, or other affective factors such as motivation. They were also young adults attending the same program at university with very similar cognitive abilities. This homogeneity might have been observed in the cognitive abilities of LA and WM in this study. Similarly, a considerable amount of the published research on LA and WM studies present data from young adults in university contexts, mostly due to the facility to reach and recruit them for research studies. Representing more variation in these factors could provide different pictures regarding the effects of L3 proficiency in LA and WM abilities. The implementation of convenience sampling could have attracted more motivated participants about their L3 experience in a period deeply affected by the Covid-19 pandemic, which has not been specifically investigated. Although assumed generalizable to a large population and not cause any systematic effects regarding the test variables in the current study, the pandemic effects might have influenced the sampling process via the emotional or physical conditions of the participants.

Future research could focus on more differentiated participant groups in the factors mentioned. A more experimental approach in the implementation and instruction conditions could also provide different views of the LA and WM constructs. A wider use of the new version of the LLAMA in particular would

contribute to its generalizability and validity across learner populations. The possible intervening testing conditions such as the instructions in this test could be investigated in more detail in wider research contexts.

APPENDIX A

BACKGROUND QUESTIONNAIRE

Bu çalışma, Boğaziçi Üniversitesi İngiliz Dili Eğitimi Programında bir doktora tez araştırması olarak yürütülmektedir. Çalışmaya katılmayı kabul ediyorsanız lütfen aşağıdaki bilgileri doldurunuz. Katılımınız için teşekkürler.

Adınız-Soyadınız:_____ *İmza:*_____ *Tarih:*_____

I. KİŞİSEL BİLGİLER (Kesinlikle Gizli kalacaktır) E-posta adresiniz: _____

Yaşınız:_____Doğum Yeriniz (Ülke):_____

II. EĞİTİM BİLGİLERİ

İlk ve ortaokuldaki eğitim dili:_____ İlk ve ortaokulda gördüğünüz yabancı dil(ler):_____
Lisede gördüğünüz eğitim dili:_____ Lisede gördüğünüz yabancı dil(ler):_____
(Varsa) Okulöncesi kurumlardaki eğitim dili:_____Okulöncesi kurumda gördüğünüz yabancı dil(ler):_____

III. DİL GEÇMİŞİ

Anadiliniz/veya dilleriniz: _____

Ailenizde konuşulan dil veya diller?:_____

Anadiliniz/dilleriniz dışında öğrendiğiniz tüm dilleri sırası ve öğrenmeye başladığınız yaşı ve yeri sırasıyla belirtiniz:

D2:_____ (Yaş:_____) (Nerede:_____)

D3:_____ (Yaş:_____) (Nerede:_____)

D4: _____ (Yaş:_____) (Nerede:_____)

Bu dilleri toplam ne kadar süredir öğreniyorsunuz? (her bir dil için ayrıca belirtiniz, *örneğin: Almanca: 8 aydır*)

D2: _____D3:_____D4:_____

Daha önce hiç İngilizce konuşulan bir ülkede kaldınız/yaşadınız mı? Cevabınız evet ise, toplam ne kadar süre ve nerede kaldınız/yaşadınız? (*örneğin:1 yıl 2 ay*)_____

Daha önce hiç üçüncü dilinizin konuşulduğu bir ülkede kaldınız/yaşadınız mı? Cevabınız evet ise, toplam ne kadar süre ve nerede kaldınız/yaşadınız? (*örneğin:1 yıl 2 ay*) _____

Aşağıdaki ortamlarda kullandığınız dilleri ve sıklıklarını bir sayı seçerek belirtiniz. Herhangi bir konumdaki dilin birden fazla olduğu durumda seçtiğiniz sayının yanında dili de belirtiniz.

Diller	Evde	Okulda	Sosyal ortamlarda (örneğin: ailede, arkadaşlarla, vb.)	Sosyal medyada
Anadil(ler)_____	1 2 3 4 5 6 1: Hiç Kullanmıyorum 6: Çok sık kullanıyorum	1 2 3 4 5 6 1: Hiç Kullanmıyorum 6: Çok sık kullanıyorum	1 2 3 4 5 6 1: Hiç Kullanmıyorum 6: Çok sık kullanıyorum	1 2 3 4 5 6 1: Hiç Kullanmıyorum 6: Çok sık kullanıyorum
İkinci dil(ler):_____	1 2 3 4 5 6 1: Hiç Kullanmıyorum 6: Çok sık kullanıyorum	1 2 3 4 5 6 1: Hiç Kullanmıyorum 6: Çok sık kullanıyorum	1 2 3 4 5 6 1: Hiç Kullanmıyorum 6: Çok sık kullanıyorum	1 2 3 4 5 6 1: Hiç Kullanmıyorum 6: Çok sık kullanıyorum
Üçüncü dil(ler)_____	1 2 3 4 5 6 1: Hiç Kullanmıyorum 6: Çok sık kullanıyorum	1 2 3 4 5 6 1: Hiç Kullanmıyorum 6: Çok sık kullanıyorum	1 2 3 4 5 6 1: Hiç Kullanmıyorum 6: Çok sık kullanıyorum	1 2 3 4 5 6 1: Hiç Kullanmıyorum 6: Çok sık kullanıyorum

III. DİL YETİSİ

Daha önce anadiliniz/dilleriniz dışındaki dillerinizde aldığınız düzey belirleme sınavlarındaki sonuçlarınızı ve hangi sınav/ları aldığınızı belirtiniz:

D2: _____ (Sınav:_____) (Düzey:_____)

D3: _____ (Sınav:_____) (Düzey:_____)

D4: _____ (Sınav:_____) (Düzey:_____)

Aşağıdaki alanlarda **anadiliniz dışındaki** dillerinizdeki seviyeleriniz "kendiniz" nasıl değerlendiriyorsunuz?

		Başlangıç	Orta	İleri	Anadil gibi
Okuma	D2:				
	D3:				
Yazma	D2:				
	D3:				
Konuşma	D2:				
	D3:				
Dinleme	D2:				
	D3:				
Genel yeti düzeyi	D2:				
	D3:				

APPENDIX B

ETHICS COMMITTEE APPROVAL

T.C.
BOĞAZİÇİ ÜNİVERSİTESİ
Sosyal ve Beşeri Bilimler Yüksek Lisans ve Doktora Tezleri Etik İnceleme Komisyonu

Sayı: 2019-15-A

7 Mart 2019

Elifcan Öztekin
Yabancı Diller Eğitimi

Sayın Araştırmacı,

"Üçüncü Dil Öğreniminde Dil Yeteneği ve İşler Bellek (Language Aptitude and Working Memory in Third Language Learning)" başlıklı projeniz ile ilgili olarak yaptığınız SBB-EAK 2019/19 sayılı başvuru komisyonumuz tarafından 7 Mart 2019 tarihli toplantıda incelenmiş ve uygun bulunmuştur.

Dr. Öğr. Üyesi İnci Ayhan

Prof. Dr. Feyza Çorapçı

Doç. Dr. Mehmet Yiğit Gürdal

Doç. Dr. Ebru Kaya

Dr. Öğr. Üyesi Şebnem Yalçın

APPENDIX C

DESCRIPTIVE STATISTICS DIVIDED BY L3 EXPERIENCE

	N	*M*	*SD*	*Min*	*Max*	*Skew.*	*Kurt.*
LLAMA B (Possible max. = 20)							
L2-only	23	9.78	3.98	3	18	.58	-.08
Elementary L3	42	8.9	4.03	3	18	.60	-.32
Intermediate L3	24	10.33	5.07	0	18	-.39	-.60
Upper int. /advanced L3	19	8.05	4.64	2	19	.89	.32
LLAMA D (Possible max. = 20)							
L2-only	23	10.17	2.92	3	15	-.91	.65
Elementary L3	40	11.2	2.68	3	18	-.31	1.77
Intermediate L3	24	13.17	2.44	9	17	-.19	-.93
Upper int. /advanced L3	19	10.74	2.21	7	14	.03	-.87
LLAMA E (Possible max. = 20)							
L2-only	23	11.3	4.99	1	20	-.36	-.60
Elementary L3	42	10.29	5.06	0	19	-.40	-.54
Intermediate L3	24	12.25	5.24	2	19	-.61	-.97
Upper int. /advanced L3	19	12.21	5.56	2	19	-.71	-.93
LLAMA F (Possible max. = 20)							
L2-only	23	11.61	3.82	4	18	-.07	-.78
Elementary L3	42	11.93	3.39	6	18	.02	-.81
Intermediate L3	24	12.83	3.38	7	19	.13	-.97
Upper int. /advanced L3	19	12.11	4.15	5	19	.18	-.45
OSpan (Possible max. = 25)							
L2-only	23	21.08	3.86	13	25	-.98	-.18
Elementary L3	41	20.54	4.11	8	25	-1.21	1.27
Intermediate L3	24	21.20	3.37	13	25	-.84	.13
Upper int. /advanced L3	19	19.63	5.16	7	25	-1.17	1.1
RotSpan (Possible max. = 14)							
L2-only	22	9.09	2.70	5	14	-.12	-1.15
Elementary L3	40	8.9	3.51	0	14	-.65	-.14
Intermediate L3	24	8.42	3.06	3	13	-.35	-1.12

Upper int. /advanced L3	19	9.63	2.19	6	14	.32	-.66
SymSpan (Possible max. = 42)							
L2-only	23	29.35	9.50	0	41	-1.4	2.92
Elementary L3	41	29.05	8.16	7	42	-.82	.65
Intermediate L3	24	28.79	5.79	21	41	.64	.92
Upper int. /advanced L3	19	27.74	7.54	13	39	-.22	-1.08
Digit span (Possible max. = 9)							
L2-only	23	7.22	1.16	5	9	-.09	-.32
Elementary L3	42	7.23	.87	4	9	-.28	1.25
Intermediate L3	24	7.17	.91	6	9	-.00	-1.14
Upper int. /advanced L3	19	6.74	1.09	4	8	-.84	.67
L3 reading comp. percentage							
Elementary L3	6	57.78	16.58	36.67	74	-.62	-1.93
Intermediate L3	16	53.23	15.8	30	84	.80	-.33
Upper int. /advanced L3	12	66.78	10.95	50	80	-.03	-1.57
L3 listening comp. percentage							
Elementary L3	6	40.56	6.21	32	50	.18	.22
Intermediate L3	16	55.04	15.8	26.67	90	.36	.54
Upper int. /advanced L3	12	64.61	11.38	50	90	.98	.74

REFERENCES

Abrahamsson, N. & Hyltenstam, K. (2008). The robustness of aptitude effects in nearnative second language acquisition. *Studies in Second Language Acquisition 30*(4), 481–509.

Alptekin, C., & Erçetin, G. (2011). Effects of working memory capacity and content familiarity on literal and inferential comprehension in L2 reading. *Tesol Quarterly*, *45*(2), 235-266.

Amaro, J. C. (2017). L3 morphosyntactic effects on L1 vs. L2 systems. In T. Angelovska & A. Hahn (Eds.), *L3 syntactic transfer: Models, new developments and implications*, Vol. 5, (pp. 173-194), Amsterdam: John Benjamins Publishing.

Artieda, G., & Muñoz, C. (2016). The LLAMA tests and the underlying structure of language aptitude at two levels of foreign language proficiency. *Learning and Individual Differences*, *50*, 42-48.

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

Baddeley, A. D., & Hitch, G. (1974). Working memory. In G. A. Flower (Ed.), *Recent advances in learning and motivation, vol. 8*. (pp. 47-90). New York: Academic Press.

Baddeley, A., Gathercole, S., & Papagno, C. (1998). The phonological loop as a language learning device. *Psychological Review 105*(1), 158-173.

Bardel, C., & Falk, Y. (2007). The role of the second language in third language acquisition: The case of Germanic syntax. *Second Language Research*, *23*(4), 459-484.

Bialystok, E. (2009). Bilingualism: The good, the bad, and the indifferent. *Bilingualism: Language and Cognition, 12*(1), 3–11.

Bialystok, E. (2011). Reshaping the mind: the benefits of bilingualism. *Canadian Journal of Experimental Psychology, 65*(4), 229-235.

Bley-Vroman, R. (1988). The fundamental character of foreign language learning. In W. Rutherford & M. Sharwood Smith (Eds.), *Grammar and second language teaching: A book of readings* (pp. 19–30). Rowley, MA: Newbury House.

Bokander, L., & Bylund, E. (2020). Probing the internal validity of the LLAMA language aptitude tests. *Language Learning*, *70*(1), 11-47.

Bouffier, M., Barbu, C., & Majerus, S. (2020). Verbal working memory but not attention is related to language proficiency: Evidence from multilingual speakers. *Psychologica Belgica, 60*(1), 270–293.

Carroll, J. B. (1962). The prediction of success in intensive foreign language training. In R. Glaser (Ed.*). Training research and education*, (pp. 87-136). Pittsburgh, PA: University of Pittsburgh Press.

Carroll, J. B. (1981). Twenty-five years of research on foreign language aptitude. In K. C. Diller (Ed.), *Individual differences and universals in language learning aptitude,* (pp. 83-118). Rowley, MA: Newbury House.

Carroll, J. B. (1990). Cognitive abilities in foreign language aptitude: Then and now. In C. Stansfield, & T. Parry (Eds.), *Language aptitude reconsidered. Language in education: Theory and practice*, (pp. 11-29). Englewoods Cliffs NJ: Prentice Halls Regents.

Carroll, J. B., & Sapon, S. M. (1959). *Modern Language Aptitude Test*. San Antonio, TX, US: Psychological Corporation.

Cockcroft, K. (2022): Working memory in multilinguals: factors that assist and impede. *International Journal of Multilingualism*, DOI: 10.1080/14790718.2022.2085710

Cockcroft, K., Wigdorowitz, M., & Liversage, L. (2019). A multilingual advantage in the components of working memory. *Bilingualism: Language and Cognition, 22*(1), 15–29

Costello, A. B., & Osborne, J. (2005). Best practices in exploratory factor analysis: Four recommendations for getting the most from your analysis. *Practical Assessment, Research, and Evaluation*, *10*(1), 7.

Çeçen, S., & Erçetin, G. (2016). Interrelationships among L2 linguistic knowledge, working memory functions, and L2 reading. In G. Granena, D. O. Jackson, & Y. Yılmaz (Eds.). *Cognitive individual differences in second language processing and acquisition*, (pp. 223-248). Amsterdam, Philadelphia: John Benjamins.

Daneman, M., & Carpenter, P. A. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning and Verbal Behavior, 19*(4), 450-466.

Daneman, M., & Merikle, P. M. (1996). Working memory and language comprehension: A meta-analysis. *Psychonomic Bulletin & Review, 3*(4), 422-433.

Dardon, D. E., & Jeong, H. (2021). Working Memory Trumps Language Aptitude in Learning Semantic-Based Linguistic Category Rules. *Studies in Language Sciences*, *19*, 77-84.

Dehn, M. (2008). *Working memory and academic learning: Assessment and intervention*. Hoboken, NJ: Wiley.

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

DeKeyser, R. (2000). The robustness of critical period effects in second language acquisition. *Studies in Second Language Acquisition 22*(04), 499–533.

DeKeyser, R. (2019). The future of language aptitude research. In Z. Wen, P. Skehan, A. Biedroń, & R. Sparks (Eds.). *Language aptitude: Advancing theory, testing, research and practice*, (pp. 317-329). New York, Oxon: Routledge Publishing.

DeKeyser, R., Alfi-Shabtay, I., & Ravid, D. (2010). Cross-linguistic evidence for the nature of age effects in second language acquisition. *Applied Psycholinguistics*, *31*(3), 413-438.

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

Ellis, R. 2004. 'Individual differences in second language learning' in A. Davies and C. Elder (Eds.). *The handbook of applied linguistics*, (pp. 525-552). Oxford: Blackwell.

Engle, R. W. (2002). Working memory capacity as executive attention. *Current Directions in Psychological Science, 11*(1), 19-23.

Engle, R. W., & Kane, M. J. (2004). Executive attention, working memory capacity, and a two-factor theory of cognitive control. In B. H. Ross (Ed.), *Psychology of learning and motivation, 44*, (pp. 145-200). Amsterdam: Elsevier.

Engle, R. W., Tuholski, S. W., Laughlin, J. E., & Conway, A. R. (1999). Working memory, short-term memory, and general fluid intelligence: a latent-variable approach. *Journal of Experimental Psychology: General, 128*(3), 309-331.

Erçetin, G., Babür, N., & Haznedar, B. (2014). *Unpublished academic achievement test battery in Turkish: Reading comprehension, listening comprehension, vocabulary, verbal memory*. Strengthening Special Education Project funded by Turkish Ministry of Education and EU. Project Code: EuropeAid/133092/C/SER/BY.

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

Espi-Sanchis, G., & Cockcroft, K. (2022). Working memory and multilingualism: balanced language proficiency predicts verbal working memory. *International Journal of Bilingual Education and Bilingualism*, *25*(8), 2976–2990.

Falk, Y., & Bardel, C. (2011). Object pronouns in German L3 syntax: Evidence for the L2 status factor. *Second Language Research, 27*(1), 59-82.

Field, A. (2013). *Discovering statistics using SPSS* (3rd ed.). SAGE Publications.

Foster, J. L., Shipstead, Z., Harrison, T. L., Hicks, K. L., Redick, T. S., & Engle, R. W. (2015). Shortened complex span tasks can reliably measure working memory capacity. *Memory & Cognition, 43*(2), 226-236.

Friedman, N. P., & Miyake, A. (2004). The reading span test and its predictive power for reading comprehension ability. *Journal of Memory and Language, 51*(1), 136-158.

Granena, G. (2013a). Cognitive aptitudes for second language learning and the LLAMA language aptitude test. In G. Granena & M. Long (Eds.), *Sensitive periods, language aptitude, and ultimate L2 attainment* (pp.105-129). Amsterdam, Philadelphia: John Benjamins Publishing Company.

Granena, G. (2013b). Reexamining the robustness of aptitude in second language acquisition. In G. Granena & M. Long (Eds.), *Sensitive periods, language aptitude, and ultimate L2 attainment* (pp. 179-204). Amsterdam, Philadelphia: John Benjamins Publishing Company.

Granena, G. (2014). Language aptitude and long-term achievement in early childhood L2 learners. *Applied Linguistics, 35*(4), 483-503.

Granena, G. (2019). Cognitive aptitudes and L2 speaking proficiency: Links between LLAMA and Hi-LAB. *Studies in Second Language Acquisition, 41*(2), 313-336.

Granena, G., Jackson, D., & Yilmaz, Y. (2016). *Cognitive individual differences in second language processing and acquisition*. Amsterdam, Philadelphia: John Benjamins Publishing.

Green, D. W., & Abutalebi, J. (2013). Language control in bilinguals: The adaptive control hypothesis. *Journal of Cognitive Psychology, 25*(5), 515-530.

Grigorenko, E. L., Sternberg, R. J., & Ehrman, M. E. (2000). A theory-based approach to the measurement of foreign language learning ability: The Canal-F theory and test. *The Modern Language Journal, 84*(3), 390-405.

Guðmundsdóttir, M. D., & Lesk, V. E. (2019). Does the bilingual advantage extend to trilingualism?. *International Journal of Multilingualism, 16*(4), 549-562.

Harrington, M., & Sawyer, M. (1992). L2 working memory capacity and L2 reading skill. *Studies in Second Language Acquisition, 14*(1), 25-38.

Herdina, P., & Jessner, U. (2002). *A dynamic model of multilingualism: Perspectives of change in psycholinguistics* (Vol. 121). Clevedon: Multilingual Matters.

Huang, T., Loerts, H., & Steinkrauss, R. (2022). The impact of second-and third-language learning on language aptitude and working memory. *International Journal of Bilingual Education and Bilingualism, 25*(2), 522-538.

Huang, Y., Wang, M., & Rao, H. (2021): Working memory as a partial mediator of the relationship between multilingualism and foreign language aptitude. *International Journal of Multilingualism*. DOI: 10.1080/14790718.2021.1993860

Hummel, K. M. (2009). Aptitude, phonological memory, and second language proficiency in nonnovice adult learners. *Applied Psycholinguistics, 30*(2), 225-249.

Iizuka, T., & DeKeyser, R. (2023). Scrutinizing LLAMA D as a measure of implicit learning aptitude. *Studies in Second Language Acquisition*, 1-23.

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

Kane, M. J., Bleckley, M. K., Conway, A. R., & Engle, R. W. (2001). A controlled-attention view of working-memory capacity. *Journal of Experimental Psychology: General, 130*(2), 169-183.

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), 189-217.

Kara-Duman, S., Yalçın, Ş., & Erçetin, G. (2021). Working memory and language aptitude in relation to listening strategy instruction in an instructed SLA context. *Annual Review of Applied Linguistics, 41,* 108-117.

Lasagabaster, D. (2000). Three languages and three linguistic models in the Basque educational system. In Cenoz, J., & Jessner, U. (Eds.). *English in Europe the acquisition of a third language* (pp. 179-197). Clevedon: Multilingual Matters.

Leeser, M. J. (2007). Learner-based factors in L2 reading comprehension and processing grammatical form: Topic familiarity and working memory. *Language Learning, 57*(2), 229-270.

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.

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

Linck, J. A., Hughes, M. M., Campbell, S. G., Silbert, N. H., Tare, M., Jackson, S. R., & Doughty, C. J. (2013). Hi-LAB: A new measure of aptitude for high-level language proficiency. *Language Learning, 63*(3), 530-566.

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,* 861-883.

Linck, J. A., Schwieter, J. W., & Sunderman, G. (2012). Inhibitory control predicts language switching performance in trilingual speech production. *Bilingualism: Language and Cognition, 15*(3), 651-662.

Ma, D., Yao, T., & Zhang, H. (2018). The effect of third language learning on language aptitude among English-major students in China. *Journal of Multilingual and Multicultural Development, 39*(7), 590–601.

Ma, D., Wang, X., & Gao, X. (2022). Effects of third-language learning on inhibitory control with consideration of task complexity and language proficiency. *International Journal of Bilingualism, 26*(3), 286-307.

Madrazo, A. R., & Bernardo, A. B. (2018). Measuring two types of inhibitory control in bilinguals and trilinguals: Is there a trilingual advantage?. *Psychological Studies, 63*(1), 52-60.

Martin, K. I., & Ellis, N. C. (2012). The roles of phonological short-term memory and working memory in L2 grammar and vocabulary learning. *Studies in Second Language Acquisition, 34*, 379-413.

Matsunaga, M. (2010). How to factor-analyze your data right: Do's, don'ts, and how-to's. *International Journal of Psychological Research, 3*(1), 97-110.

Meara, P. & Rogers, V. E. (2019). *The LLAMA Tests v3*. Cardiff: Lognostics.

Meara, P. (2005). *LLAMA language aptitude tests*. Swansea: Lognostics.

Mitchell, A. E., Jarvis, S., O'Malley, M., & Konstantinova, I. (2015). 15. Working memory measures and L2 proficiency. In Z. Wen, M. Bota, & A. Mcneill (Eds.). *Working memory in second language acquisition and processing*, (pp. 270-283). Bristol, New York, Ontario: Multilingual Matters.

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

Osaka, M., & Osaka, N. (1992). Language-independent working memory as measured by Japanese and English reading span tests. *Bulletin of the Psychonomic Society, 30*(4), 287-289.

Osaka, M., Osaka, N., & Groner, R. (1993). Language-independent working memory: Evidence from German and French reading span tests. *Bulletin of the Psychonomic Society, 31*(2), 117-118.

Paradis, M. (2009). Declarative and procedural determinants of second languages. *Declarative and procedural determinants of second languages*. Amsterdam, Philadelphia: John Benjamins.

Payne, T. W., Kalibatseva, Z., & Jungers, M. K. (2009). Does domain experience compensate for working memory capacity in second language reading comprehension?. *Learning and Individual Differences, 19*(1), 119-123.

Petersen, C. R., & Al-Haik, A. R. (1976). The development of the defense language aptitude battery (DLAB). *Educational and Psychological Measurement, 36*(2), 369-380.

Pimsleur, P. (1966). *Pimsleur language aptitude battery (PLAB)*. New York: Harcourt, Brace, Jovanovich.

Poarch, G. J., & Van Hell, J. G. (2012). Executive functions and inhibitory control in multilingual children: Evidence from second-language learners, bilinguals, and trilinguals. *Journal of Experimental Child Psychology, 113*(4), 535-551.

Redick, T. S., Broadway, J. M., Meier, M. E., Kuriakose, P. S., Unsworth, N., Kane, M. J., & Engle, R. W. (2012). Measuring working memory capacity with automated complex span tasks. *European Journal of Psychological Assessment, 28*(3), 164-171.

Robinson, P. (1997). Individual differences and the fundamental similarity of implicit and explicit adult second language learning. *Language Learning, 47*(1), 45-99.

Robinson, Peter (2002). Effects of individual differences in intelligence, aptitude and working memory on adult incidental SLA: A replication and extension of Reber, Walkenfeld and Hernstadt, 1991. In P. Robinson (Ed.), *Individual differences and instructed language learning*, (pp. 211–266), Amsterdam: John Benjamins.

Rogers, V., Meara, P., Barnett-Legh, T., Curry, C., & Davie, E. (2017). Examining the LLAMA aptitude tests. *Journal of the European Second Language Association, 1*(1), 49–60.

Sáfár, A., & Kormos, J. (2008). Revisiting problems with foreign language aptitude. *IRAL- International Review of Applied Linguistics in Language Teaching, 46*(2), 113-136.

Saito, K. (2017). Effects of sound, vocabulary, and grammar learning aptitude on adult second language speech attainment in foreign language classrooms. *Language Learning, 67*(3), 665-693.

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.

Sawyer, M., & Ranta, L. (2001). Aptitude, individual differences, and instructional design. In P. Robinson (Ed.), *Cognition and second language instruction*, (pp. 319353). Cambridge: Cambridge University Press.

Schrank, F. A., McGrew, K. S., & Woodcock, R. W. (2001). *Technical abstract* (Woodcock-Johnson III Assessment Service Bulletin No. 2). Itasca, IL: Riverside Publishing.

Schwieter, J. W., & Sunderman, G. (2011). Inhibitory control processes and lexical access in trilingual speech production. *Linguistic Approaches to Bilingualism, 1*(4), 391-412.

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.

Service, E., Simola, M., Metsänheimo, O., & Maury, S. (2002). Bilingual working memory span is affected by language skill. *European Journal of Cognitive Psychology, 14*(3), 383-408.

Shah, P. & Miyake, A. (1996). Models of working memory. In A. Miyake & P. Shah (Eds.), *Models of working memory: Mechanisms of active maintenance and executive control*, (pp. 1-27). New York: Cambridge University Press.

Skehan, P. (2002). Theorising and updating aptitude. In P. Robinson (Ed.). *Individual differences and instructed language learning*, (pp. 69-95). Amsterdam/Philadelphia: John Benjamins.

Skehan, P. (2016). Foreign language aptitude, acquistional sequences, and psycholinguistic processes. In G. Granena, D. O. Jackson, & Y. Yılmaz (Eds.). *Cognitive individual differences in second language processing and acquisition*, (pp. 17-40). Amsterdam, Philadelphia: John Benjamins Publishing.

Skehan, P. (2019). Language aptitude implicates language and cognitive skills. In Z. Wen, P. Skehan, A. Biedroń, S. Li, & R. Sparks (Eds.), *Language aptitude: Advancing theory, testing, research and practice* (pp. 56–77). Routledge.

Sok, S., & Shin, H. W. (2022). Investigating the role of cognitive variables in second language learners' listening comprehension: Aptitude and metacognitive awareness. *International Journal of Listening, 36*(2), 138-151.

Sparks, R. L., Humbach, N., Patton, J. O. N., & Ganschow, L. (2011). Subcomponents of second-language aptitude and second-language proficiency. *The Modern Language Journal, 95*(2), 253-273.

Suzuki, Y. (2021). Probing the construct validity of LLAMA_D as a measure of implicit learning aptitude: Incidental instructions, confidence ratings, and reaction time. *Studies in Second Language Acquisition, 43*(3), 663-676.

Thompson, A. S. (2013). The interface of language aptitude and multilingualism: Reconsidering the bilingual/multilingual dichotomy. *The Modern Language Journal, 97*(3), 685–701.

Turner, M. L., & Engle, R. W. (1986). Working memory capacity. In *Proceedings of the human factors society 30th annual meeting*, (pp. 1273-1277). Sage CA: Los Angeles, CA: SAGE Publications.

Turner, M. L., & Engle, R. W. (1989). Is working memory capacity task dependent?. *Journal of Memory and Language, 28*(2), 127-154.

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). Washington, DC: Georgetown University Press.

Unsworth, N., & Engle, R. W. (2005). Individual differences in working memory capacity and learning: Evidence from the serial reaction time task. *Memory & Cognition, 33*(2), 213-220.

Van den Noort, M. W., Bosch, P., & Hugdahl, K. (2006). Foreign language proficiency and working memory capacity. *European Psychologist, 11*(4), 289-296.

Was, C. A., & Woltz, D. J. (2007). Reexamining the relationship between working memory and comprehension: The role of available long-term memory. *Journal of Memory and Language, 56*(1), 86-102.

Waters, G. S., & Caplan, D. (1996). The measurement of verbal working memory capacity and its relation to reading comprehension. *The Quarterly Journal of Experimental Psychology Section A, 49*(1), 51-79.

Wen, Z., & Skehan, P. (2011). A new perspective on foreign language aptitude research: building and supporting a case for" working memory as language aptitude". *Ilha do Desterro A Journal of English Language, Literatures in English and Cultural Studies*, (60), 015-044.

Wen, Z. & Skehan, P. (2021). Stages of acquisition and the P/E model of working memory: Complementary or contrasting approaches to foreign language aptitude?. *Annual Review of Applied Linguistics, 41*. 6-24.

Wen, Z. (2011). Foreign language aptitude. *ELT Journal, 66*(2), 233-235.

Wen, Z. (2015). Working memory in second language acquisition and processing: The phonological/executive model. In Z. Wen, M. Bota, & A. Mcneill (Eds.). *Working memory in second language acquisition and processing*, (pp. 41-62). Bristol, New York, Ontario: Multilingual Matters.

Wen, Z. E., Biedroń, A., & Skehan, P. (2017). Foreign language aptitude theory: Yesterday, today and tomorrow. *Language Teaching, 50*(1), 1-31.

Williams, J. N., & Lovatt, P. (2003). Phonological memory and rule learning. *Language Learning, 53*(1), 67-121.

Winke, P. (2013). An investigation into second language aptitude for advanced Chinese language learning. *The Modern Language Journal, 97*(1), 109–130. https://doi.org/10.1111/J.1540-4781.2013.01428.X

Xiang, H., Dediu, D., Roberts, L., Oort, E. van, Norris, D. G., & Hagoort, P. (2012). The structural connectivity underpinning language aptitude, working memory, and IQ in the Perisylvian language network. *Language Learning, 62*(SUPPL. 2), 110–130. https://doi.org/10.1111/J.1467-9922.2012.00708.X

Yalçın, Ş., Çeçen, S., & Erçetin, G. (2016). The relationship between aptitude and working memory: An instructed SLA context. *Language Awareness, 25*(1-2), 144-158.

Yalçın, Ş., & Spada, N. (2016). Language aptitude and grammatical difficulty: An EFL classroom-based study. *Studies in Second Language Acquisition, 38*(2), 239-263.