Language learning aptitude and working memory in L2 acquisition: The role of proficiency and structure difficulty

Renato Pavlekovic

A thesis submitted for the degree of

Doctor of Philosophy in Applied Linguistics

Department of Language and Linguistics

University of Essex

December 2024

Statement of Authorship

The Candidate confirms that the presented work is original and has not been submitted for examination elsewhere. The Candidate further certifies that this thesis does not, to the best of their knowledge, infringe on anyone's copyright or violate any proprietary rights, and that material sourced from other authors has been cited and credited accordingly. Copies of such material have been made openly available in the Appendices.

The Candidate acknowledges the contribution of Dr Karen Roehr-Brackin as a co-author of Chapters 4 and 5, and certifies authorship and publication status of the Chapters presented in this thesis as follows:

- Chapter 4 (Article 1): authored by the Candidate, co-authored by Dr Karen Roehr-Brackin. Published in edited form as Pavlekovic, R., & Roehr-Brackin, K. (2024). Aptitude for Explicit and Implicit Learning. *Journal of the European Second Language Association*, 8(1), 117. https://doi.org/10.22599/jesla.117

- Chapter 5 (Article 2): authored by the Candidate, co-authored by Dr Karen Roehr-Brackin. At the time of submission of this thesis, submitted and under review with *Studies in Second Language Acquisition*.

- Chapter 6 (Article 3): authored by the Candidate, supervised by Dr Karen Roehr-Brackin. At the time of submission of this thesis, submitted and under review with *Language Learning*.

- Chapter 7: authored by the Candidate, supervised by Dr Karen Roehr-Brackin. Not awaiting submission.

- Chapter 8: authored by the Candidate, supervised by Dr Karen Roehr-Brackin. Not awaiting submission.

Abstract

Research on language aptitude has empirically demonstrated its facilitative role in second language learning (L2) outcomes. However, relatively few studies have examined how these effects evolve with increasing L2 proficiency, and even fewer have explored the role of linguistic structure difficulty as a moderating factor. This thesis investigates whether the influence of language aptitude remains stable or changes in learners at different proficiency levels. It also examines the impact of linguistic structure difficulty and addresses the long-standing question of the interface between implicit and explicit knowledge by analysing the relationship between aptitude and knowledge of targeted morphosyntactic structures.

Eighty-six L1 Croatian learners of English completed the LLAMA aptitude battery and a serial reaction time task to assess aptitude for explicit and implicit learning, as well as forward digit span and operation span tasks to assess phonological and executive working memory, respectively. Reading and listening proficiency was assessed via the Oxford Placement Test, while speaking proficiency was operationalized as complexity, accuracy and fluency measured in an oral production task. L2 knowledge of selected English morphosyntactic structures of varying difficulty – articles, passive, and the simple past – was measured using a self-paced reading task, elicited imitation, and gap-fill tests.

The thesis explores the componential structure of language aptitude and working memory, concluding that both implicit and explicit aptitude are multicomponential constructs. Notably, implicit aptitude is shown to differ from explicit aptitude with its components pulling in opposite directions and is reconceptualized as a cognitive proclivity. The findings reveal that the facilitative effects of language aptitude shift with proficiency: explicit aptitude is more relevant at lower proficiency levels, while implicit aptitude becomes increasingly important at advanced levels. Additionally, the study demonstrates a bidirectional interface between implicit and explicit aptitude and L2 knowledge. This means that implicitly learned knowledge can lead to explicit knowledge, while explicitly learned knowledge can facilitate the development of implicit knowledge. The moderating role of structure difficulty is confirmed, highlighting its influence on the effects of working memory and explicit aptitude, though not implicit aptitude.

In sum, the findings confirm facilitative effects of both explicit and implicit aptitude. Moreover, they position aptitude as a dynamic concept. The evidence for a dynamic interface between explicit and implicit knowledge highlights the interplay between learning processes, suggesting that the traditional question of whether the interface exists may be outdated.

Acknowledgments

To my love, Katarina, thank you for standing by me through every step of this journey. Your unwavering support and encouragement have been my greatest source of strength. You have been my rock, my anchor, and my biggest cheerleader. This achievement would not have been possible without you – it is as much yours as it is mine.

To my beloved family: my mom, Silvija, and my brother, Ervin, thank you for always believing in me and reminding me that I can do whatever I put my mind to. Your love and understanding have been my light in the darkness. To my grandparents, Štefica and Vlado, thank you for all your support and encouragement. Your eternal positivity and our video calls made this experience so much more enjoyable. To the rest of my family – Dijana, Neven, Ivan, and Gea – thank you for always believing in me and being my steadfast support.

A special thanks to Dr Karen Roehr-Brackin from the bottom of my heart. Your advice and support were always precisely what I needed, exactly when I needed them to get through this process. You consistently encouraged my curiosity and independence while providing expert guidance and reassurance. I will always admire your research ingenuity and scientific rigor. This has been an amazing experience that I will forever treasure.

Profound gratitude goes to the members of my supervisory panel: Dr Laurel Lawyer, Dr Adela Gánem-Gutiérrez, and Dr Kyle Jerro, for willingly sharing your expertise and placing your faith in me and my work.

I would also like to express my gratitude to the members of the Applied and Experimental Linguistics Research Group at the University of Essex for your valuable feedback on the analyses and interpretation of the results of this study. Thank you all!

Table of Contents

1. Introduction

Language learning is an intrinsically fascinating and personally rewarding experience that enriches individuals and contributes to societal development. It is cognitively stimulating, with the benefits of learning a new language and bilingualism often compared to those of other demanding activities, such as juggling, using novel tools, and navigation (Draganski et al., 2004; Maguire et al., 2000; Quallo et al., 2009; Taubert et al., 2010).

Despite these benefits, learning a second language (L2) is often seen as a difficult, protracted, and challenging process, deterring many individuals from starting. A common misconception is that achieving advanced proficiency in adulthood is all but impossible, sometimes attributed to a so-called “talent for languages” that only a fortunate few possess. Since the 1950s, researchers in second language acquisition (SLA) have investigated this idea through studies on language aptitude, which refers to the cognitive and perceptual abilities that enable to learn languages effortlessly and facilitate high achievement in language learning (Granena, 2020).

Early research in this domain primarily focused on developing reliable measures of aptitude to identify learners with high potential, based on the assumption that aptitude was a fixed set of cognitive abilities (Carroll, 1962; Carroll & Sapon, 1959). Over time, however, a broader understanding of aptitude has emerged, showing that it can serve purposes beyond merely identifying “talented” learners.

A key paradigm shift occurred when researchers moved from using aptitude tests as predictive tools to exploring the explanatory potential of aptitude. This reframing positioned aptitude as a dynamic factor, enabling investigations into its interaction with other variables. Empirical studies investigating the effects of language aptitude have provided ample evidence of the benefits of high aptitude levels (Li, 2015), but they also reveal that these benefits do not always play a role. For example, some studies have found that learners in the early stages of language acquisition benefit more from higher levels of aptitude than those at intermediate or advanced levels (Artieda & Muñoz, 2016). This suggests that while high language aptitude can be advantageous, it is not always essential, and learners with lower aptitude are not necessarily always at a disadvantage. This has informed theoretical frameworks such as Skehan's developmental model (Skehan, 2002, 2012, 2016) and Robinson's aptitude complexes model (Robinson, 2005, 2007, 2012) which relate the concept of language aptitude to the learning process while also acknowledging the influence of other factors on the effects of aptitude, such as task complexity or the type of instruction.

Ever since dual-system theories of human cognition were introduced to second language acquisition (SLA), particularly the distinction between explicit and implicit knowledge (Krashen, 1981, 1982, 1985), researchers have agreed that humans learn languages both consciously and intentionally, such as through formal instruction in a classroom setting, and unintentionally and incidentally, for example, by living in a country where the language is spoken and improving proficiency without formal lessons or deliberate efforts to learn grammar.

This duality sparked debate among linguists about the "best" way to learn a language and achieve high proficiency. While both types of learning were shown to be effective, researchers were particularly interested in understanding how to acquire unconscious, implicit knowledge, which is highly automatic and used in contexts like speaking. It also became evident that some individuals, even after prolonged immersion in a target-language environment, failed to achieve high levels of proficiency.

This observation led researchers to explore the relationship between implicit (unconscious) and explicit (conscious) knowledge. They questioned whether it was possible to develop implicit, automatic knowledge through intentional learning or if learning unintentionally could foster explicit awareness of language rules. Although few studies have addressed this issue empirically, initial findings suggest that the relationship is bidirectional: intentional learning can lead to implicit knowledge (Suzuki & DeKeyser, 2017), while incidental learning can result in the emergence of explicit awareness (Kim & Godfroid, 2023).

Lastly, not all aspects of a language are equally difficult, despite the fact that language learning overall is, for most people, a demanding task. Some of that difficulty can be a result of opaqueness of language, i.e. low transparency around how a linguistic feature works, or when a linguistic feature shows low systematicity, i.e. a grammar rule that has a lot of exceptions (Roehr & Gánem-Gutiérrez, 2009a). So, if language aptitude is a set of cognitive abilities that facilitate language learning, a question that presents itself is whether language aptitude is equally relevant for all grammatical structures, or whether its role depends on the difficulty of a linguistic structure. Only a handful of studies have looked at this problem in a systematic way, and those who did found initial evidence that more difficult structures often mean heavier reliance on language aptitude (Robinson, 2002; Yalçın & Spada, 2016).

In addition to the distinction between explicit and implicit knowledge, researchers also hypothesized the existence of corresponding aptitudes for explicit and implicit learning. Research in this area is still at an early stage, but initial findings suggest that these distinctions do exist (Granena, 2013a; Roehr-Brackin et al., 2023). However, no study to date has incorporated this distinction into its methodology.

Despite these advancements, fundamental questions remain unanswered. Can cognitive abilities for explicit and implicit learning be reliably differentiated? If so, which abilities belong to each construct? Furthermore, should working memory be considered part of language aptitude, or is it distinct enough to require separate measures? Finally, given the facilitative effects of cognitive aptitudes on language outcomes, how do factors such as proficiency level and linguistic structure difficulty mediate or moderate the influence of aptitude and working memory?

This project seeks to address these questions in the context of a classroom setting where language aptitude is said to predict the rate of learning (Granena, 2013c). The componential structure of language aptitude is examined by means of a comprehensive range of contemporary cognitive and linguistic measures. By operationalizing proficiency as a continuous variable, this study aims to determine where along the proficiency continuum explicit and implicit language aptitude begin to exert their effects and where their influence diminishes. Testing L2 knowledge across linguistic structures of varying difficulty will provide insights into whether, and to what extent, structure difficulty moderates the effects of language aptitude. Furthermore, examining the relationship between language aptitude and L2 knowledge will further our understanding of the interface between implicit and explicit knowledge.

2. Review of literature

2.1. Knowledge in L2

As Bialystok (1994b) notes, most L2 learners recognize that there are aspects of their L2 that they consciously learn and can access through introspection, while other aspects operate outside their conscious awareness but are still executed correctly. In psychology, dual-process or dual-system theories of higher cognition explain how thought can arise through two distinct pathways or processes – implicit and explicit. Implicit and explicit cognition are viewed as architecturally and evolutionarily distinct, involving two separate processing systems: automatic and controlled (Evans & Over, 1996; Granena, 2020; Stanovich & West, 2000). In second language acquisition, the idea of two separate learning systems – implicit, which is nonconscious and takes longer, and explicit, which is conscious and faster – aligns with established notions of implicit and explicit language learning and aptitude. Beginning with Krashen's seminal work (1981), empirical evidence in SLA has similarly pointed to the existence of two distinct types of language knowledge – implicit and explicit. The following sections will introduce the topic of implicit and explicit knowledge and learning, define current theoretical accounts of their relationship, and provide an overview of widely used and empirically validated measures for assessing implicit and explicit knowledge in adult L2 learners.

2.1.1. Implicit and explicit knowledge

2.1.1.1. Implicit knowledge and learning

Implicit learning, a term first introduced by Reber (1967), describes a process by which individuals acquire knowledge about a complex, rule-governed stimulus domain without conscious intent or awareness of what has been learned. In his seminal study, Reber exposed participants to a series of sentences that adhered to a finite-state grammar in an artificial language. During the testing phase, learners were able to correctly judge grammatical and ungrammatical sentences with 79% accuracy, despite being unable to articulate the underlying rules. This provided compelling evidence for implicit learning, demonstrating that individuals can unconsciously acquire knowledge of artificial grammar. Reber concluded that implicit learning is a "rudimentary inductive process" (Reber, 1967, p. 863), potentially intrinsic to processes like language learning.

Implicit language learning involves chunk learning rather than rule learning, where learners internalize sequences of sounds and words based on their frequency in the input. Abstractions, or rule-like connections, are then formed from these sequences (N. Ellis, 2015). Implicit learning relies on similarity-based processing and requires substantial input exposure to be effective. The product of implicit learning, implicit knowledge, is often described as tacit and intuitive, inaccessible to conscious introspection and non-verbalizable (Hulstijn, 2015; Rebuschat, 2013). Although implicit learning is a slow process, implicit knowledge is accessed rapidly and effortlessly and is said to underlie communicative competence in spontaneous comprehension and production (R. Ellis, 2005).

Reber's (1992, 1996) evolutionary perspective on implicit learning highlights why this process is of great interest to both researchers and learners. First, implicit knowledge is unconscious and inaccessible to conscious reflection. As Reber points out, consciousness is a relatively recent evolutionary development, built upon deeper and more primitive processes that operate outside of conscious awareness, such as implicit learning. Second, evolutionary theory suggests that phylogenetically older, more primitive cognitive structures are more robust and less vulnerable to disruption than newer ones. This implies that implicit processes may exhibit greater resilience to neurological and clinical disorders than explicit processes.

Most importantly, evolutionarily older cognitive functions, having been shaped by millennia of adaptation, exhibit less variation between individuals and a tighter distribution within populations than newer functions. Therefore, implicit processes are expected to show fewer individual differences compared to explicit processes. This lack of variation may explain the relatively neglected status of implicit learning in research until recent years. If implicit learning abilities are largely uniform across individuals, then efforts to capitalize on strengths or compensate for weaknesses in the context of SLA might seem futile. However, more recent evidence has revealed considerable variation across participants in standard implicit learning tasks in SLA (Kalra et al., 2019; Kaufman et al., 2010; Woltz, 2003), a point that Reber himself acknowledged (Reber & Allen, 2000). This raises important questions about the nature of implicit learning as an ability (Rebuschat, 2022), one of which is the structure of implicit learning ability and its predictive power for language learning outcomes.

Statistical learning, a process involving the extraction of distributional properties from the environment, shares striking similarities with implicit learning. Some researchers argue that statistical learning and implicit learning represent two perspectives on the same phenomenon (Perruchet, 2019; Rebuschat, 2022), while others propose combining the terms into "implicit statistical learning" (Conway & Christiansen, 2006). It is important to clarify that implicit learning is a broad construct which includes the more narrow concept of statistical learning. For instance, Bogaerts et al. (2020) suggest that statistical learning, along with sequence and procedural learning, constitute distinct forms of implicit learning, characterized by an unconscious learning process. However, these constructs often overlap in the experimental tasks designed to assess them. In the context of implicit language learning, tasks intended to probe statistical learning are frequently used to measure implicit learning ability. Consequently, and in line with the prevailing terminology in current research, the term "implicit learning" is used throughout as a synonym with statistical learning.

2.1.1.2. Explicit knowledge and learning

Explicit learning, in contrast to implicit learning, involves a conscious and deliberate process (Rebuschat, 2013) where learners actively make and test hypotheses about the language they are acquiring (N. Ellis, 2015). This process requires attentional resources for information processing and maintenance in working memory, making it cognitively demanding (Roehr-Brackin, 2015). Explicit learning is a fast and efficient process; unlike implicit learning which relies on ample input for frequency-based tallying, explicit learning can achieve high efficiency with minimal input. Ullman (2020) suggests that linguistic knowledge stored in declarative memory (i.e., explicit knowledge) can sometimes be acquired from a single exposure, provided the information is simple enough. Explicit learning facilitates the formation of form-meaning mappings and can be used in controlled production, potentially contributing to the development of more automatized knowledge (N. Ellis, 2011).

Explicit knowledge, the product of explicit learning, is conscious knowledge that can be verbalized and accessed through introspection (N. Ellis, 2015; Rebuschat, 2013). It is described *knowledge how*, encompassing knowledge of language features as well as the metalanguage needed to label these features (R. Ellis, 2005). This knowledge is typically accessed through controlled processing (R. Ellis, 2004, 2005). Roehr (2008) further differentiates explicit knowledge into two categories: metalinguistic knowledge, or learners' ability to correct, describe, and explain L2 errors, and metalanguage, the technical or semi-technical terms used to describe language, often learned through instruction (R. Ellis, 2009a).

Explicit knowledge is intimately related to explicit instruction. Its effectiveness is particularly acknowledged for linguistic forms that lack salience (Schmidt, 1990, 2001, 2007) or when L2 semantic or pragmatic concepts are mapped onto L2 forms in unfamiliar ways, necessitating additional attention (N. Ellis, 2015). A substantial body of research has demonstrated that form-focused L2 instruction leads to significant target-oriented gains, with explicit instruction proving more effective than implicit methods (R. Ellis, 2001; Norris & Ortega, 2000, 2006; Spada, 1997; Spada & Tomita, 2010)

Neurobiological models of L2 acquisition (Paradis, 2009; Ullman, 2020) explain the dual nature of L2 knowledge by associating it with distinct memory systems and brain regions. Explicit knowledge is linked to declarative memory, a domain-general system for learning facts and events, while implicit knowledge is associated with procedural memory, which is responsible for learning motor and cognitive skills and forming habits (Suzuki et al., 2023). Neurological evidence has connected declarative memory to the hippocampus and medial temporal lobe structures, showing its involvement in grammar learning when metalinguistic rules are provided (Tagarelli et al., 2019). Procedural memory, by contrast, is linked to the frontal cortical-basal ganglia system, which plays a critical role in grammar learning (Suzuki et al., 2023). Ullman (2020) further argues that the anterior caudate nucleus is crucial during the early stages of grammar learning, while the premotor cortex and inferior frontal gyrus become more involved as skills become automatized (Suzuki et al., 2023). Importantly, the Declarative-Procedural (D/P) model (Ullman, 2004, 2016, 2020) posits a competing relationship between the two memory systems. For instance, representations learned through declarative memory may inhibit analogous representations formed in procedural memory, and vice versa, depending on which system is dominant for a given representation. Preliminary evidence for this comes from a negative correlation observed between the two systems during performance on an elicited imitation task (Suzuki et al., 2023).

2.1.1.3. The role of implicit and explicit learning in SLA theories

From a usage-based perspective, most language acquisition occurs through use and implicit learning (N. Ellis, 2005). According to this view, implicit learning systems drive the process by detecting sequences based on the frequency of linguistic cues, while explicit learning aids in vocabulary acquisition and draws attention to non-salient structures, after which implicit learning processes take over.

The Interaction Hypothesis (Long, 1996) posits that optimal conditions for L2 learning happen when learners' attention is directed to linguistic forms during meaning-primary interaction while keeping the attention to form brief and ideally in the form of implicit negative feedback. Similar to the usage-based perspective, interaction hypothesis views implicit learning as the default mechanism for L2 learning, with children having access to the full potential of implicit learning mechanisms, while adult L2 learners experience gradually restricted access. In these cases, intentional or conscious learning helps establish new form-meaning connections, albeit requiring minimal attentional resources and no metalinguistic awareness.

In contrast, Skill Acquisition Theory (DeKeyser, 2015, 2020) argues that language skills develop gradually, moving from explicit to implicit knowledge through stages of declarative, procedural, and finally automatic knowledge, a process driven by repeated practice. According to this view, explicit learning is essential in the early stages (declarative stage), while implicit learning becomes relevant in later (procedural and automatic) stages.

Finally, the Declarative-Procedural (D/P) Model (Ullman, 2004, 2016, 2020) also supports the distinction between implicit and explicit learning. According to the model, declarative memory aligns with explicit aptitude, while procedural memory corresponds to implicit aptitude, although it is acknowledged that aptitude encompasses more than just memory. For example, declarative memory has been operationalized as associative and recognition memory, while within the aptitude paradigm, explicit aptitude traditionally includes associative memory, phonetic coding ability, and language-analytic ability (Li & DeKeyser, 2021). The D/P model posits that declarative memory is responsible for learning idiosyncratic information and arbitrary associations, whereas procedural memory governs grammar learning and the acquisition of sequences, rules, and patterns.

2.1.2. Relationship between implicit and explicit knowledge

The relationship between implicit and explicit knowledge has been a longstanding issue in the field, commonly referred to as the “interface debate”. This debate centres around the question of whether the two types of linguistic knowledge are interconnected, and if so, how they relate to one another. Broadly speaking, three major positions have emerged: the non-interface position, the weak interface position, and the strong interface position. More recently, theoretical re-examinations of this debate have proposed a bidirectional relationship and simultaneous existence of multiple interfaces, suggesting a more nuanced relationship between the two types of knowledge.

2.1.2.1. Non-interface

Krashen’s (1981) Monitor Model gave rise to the non-interface position. Krashen argues that implicit and explicit knowledge are distinct constructs acquired through separate mechanisms, with no interaction between them. According to this theory, language is too complex to be learned or explained explicitly. Instead, it is primarily acquired through implicit learning mechanisms, while explicit knowledge, established via explicit learning, serves only as a monitor of L2 utterances before or after they are produced by the implicit system. As Krashen notes, “error correction and explicit teaching of rules are not relevant to language acquisition” (Krashen, 1981, p. 1). Given that fluent language use is reliant on implicit knowledge and that explicit knowledge is slower to access, monitoring is ineffective in real-time communication. Consequently, the Monitor Model maintains that explicit knowledge cannot interface with implicit knowledge in any meaningful way. Thus, language learning is driven primarily by the development of implicit knowledge, with explicit knowledge playing a minimal role and having no direct influence on or interaction with implicit knowledge.

Hulstijn's position on the interface issue has shifted significantly over time. Initially advocating a strong interface (Hulstijn, 1990), he later adopted a view that aligns more closely with the non-interface position, arguing that explicit knowledge cannot transform into implicit knowledge (Hulstijn, 2005, 2015). However, unlike Krashen, Hulstijn considers explicit knowledge an essential form of knowledge, particularly in situations "where and when implicit knowledge is not (yet) available" (Hulstijn, 2015, p. 209).

Paradis (1994, 2009) approached the interface issue from a cognitive neuroscience perspective, asserting that implicit and explicit knowledge are inseparable from their respective memory systems – explicit knowledge resides in declarative memory and implicit knowledge in procedural memory, both of which are located in different brain regions. Consequently, he argues, they cannot interface.

2.1.2.2. Strong interface

In contrast, the strong interface position is rooted in Skill Acquisition Theory (DeKeyser, 2015, 2020), which is based on Anderson's (1982, 2007) cognitive skills view from cognitive psychology. Skill Acquisition Theory views adult language learning as predominantly a conscious process, starting with explicit knowledge. Learning progresses through distinct stages: beginning with the declarative stage, where learners develop factual knowledge ("knowledge that"), then moving to the procedural stage ("knowledge how"), and finally reaching the automatic stage, where knowledge becomes fluent, spontaneous, and effortless. These stages are sometimes referred to as "cognitive /associative /autonomous" when referring to the learning theory, or "presentation /practice /production" when referring to the application to L2 teaching. This framework aligns with Schmidt's Noticing Hypothesis Schmidt (1990, 2001), which posits that language learning cannot occur through "subliminal perception"

(Schmidt, 1990, p. 142); rather, it requires conscious attention. According to DeKeyser, implicit knowledge is derived from explicit knowledge through repeated practice (DeKeyser, 1998, 2015, 2020). However, he argues that explicit knowledge is not transformed into implicit knowledge, as this would imply that explicit knowledge disappears once implicit knowledge is acquired, which is not the case. Instead, DeKeyser suggests that repeated use of one memory system gradually leads to the establishment of the other, with explicit knowledge linked to declarative memory and implicit knowledge associated with procedural memory.

2.1.2.3. Weak interface

N. Ellis's (2005, 2006; 2015) interface position is embedded in the Associative-Cognitive CREED framework (N. Ellis, 2006), which posits that the cognitive processes involved in language acquisition are similar to those governing other forms of human cognition and that SLA is driven by general laws of learning, such as associative and cognitive principles. The primary distinction between learning a language and other skills lies in the unique cognitive content of language systems. According to N. Ellis, most language learning occurs implicitly through exposure and usage (N. Ellis & Wulff, 2020). As he states, "most knowledge is tacit, most learning is implicit, [and] the vast majority of our cognitive processing is unconscious" (p. 77). Implicit learning mechanisms allow learners to form generalizations and abstractions based on the distributional properties of the input. However, certain aspects of an L2 may either be perceptually non-salient or lack transparent form-meaning mappings, making them difficult to acquire implicitly. In such cases, additional attentional resources are required, and explicit learning becomes necessary. This often occurs in instructed L2 contexts, where explicit instruction aids in the development of explicit representations. Once these constructions are explicitly represented, they facilitate the initial recognition of patterns and enable further statistical tallying of usage frequency and form-meaning probabilities, contributing to implicit learning through distributional analysis. N. Ellis argues that the relationship between implicit and explicit knowledge is dynamic, with the two types of knowledge interfacing temporarily during conscious processing. This interaction leaves a lasting impact on implicit cognition, allowing explicit learning to influence implicit knowledge formation over time (N. Ellis, 2005; 2015).

2.1.2.4. Implicit-explicit and reciprocal interface

Bialystok's framework of analysis and control (1994a, 1994b, 2001), designed to address the key aspects of cognitive mechanisms that account for language acquisition and use, tacitly offers another perspective on the relationship between implicit and explicit knowledge. Rooted in cognitive psychology, this framework posits that the analysis of knowledge reflects one's ability to mentally represent increasingly explicit and complex structures in a systematic way. Control of processing refers to the ability to selectively allocate attentional resources to task-relevant aspects while inhibiting task-irrelevant information. The need for these abilities varies dynamically depending on task demands and context. Though originally emerging from research on bilingual children, the framework is applicable to L2 settings. It proposes that L2 knowledge begins implicitly, with learners gradually developing awareness of linguistic structures through the increasing analysis of knowledge. In this way, explicit knowledge emerges from implicitly accumulated linguistic knowledge.

The existence of an implicit-explicit interface has been empirically demonstrated by Kim and Godfroid (2023), the first study in an L2 research context using a natural language to find such evidence. Their longitudinal study showed that implicit knowledge at time point one influenced the development of explicit knowledge at time point two, and vice versa. This was interpreted as evidence of a bidirectional or reciprocal interface, suggesting that the relationship between the two types of knowledge is both dynamic and mutually influential.

2.1.2.5. Co-existence of interfaces

It has also been proposed that multiple interfaces between explicit and implicit knowledge can exist simultaneously (Han & Finneran, 2014). According to this view, different relationships between these types of knowledge may coexist within and across various linguistic subsystems, as well as among different learners. Han & Finneran reached this conclusion after considering the partial validity of each of the three traditional interface positions (non-interface, weak interface, strong interface), while also recognizing the counter-evidence each position fails to address.

The study focused on a range of L2 structures examined in previous research on individual learners, analysing instances where learners made morphosyntactic errors of the same type on some occasions but not others. These cases were contrasted with examples of morphosyntactic structures for which learners made consistent errors. The former was interpreted as a lack of interface, while the latter was seen as evidence of an interface.

This reasoning was applied to explain the absence of explicit knowledge of plurals and articles observed in a learner who had been living in a target-language-speaking country for over a decade. While the learner demonstrated relatively high accuracy on tasks engaging implicit knowledge, their performance was markedly lower on tasks requiring explicit knowledge of the same grammatical structures. The authors concluded that the extensive input received by the learner over many years had little impact on their explicit knowledge of plurals and articles. They argued that this scenario illustrates a lack of interface between the implicit knowledge gained through prolonged exposure and the explicit knowledge needed for error correction tasks.

In contrast, the authors also examined examples of grammatical errors that were consistently made across different tasks. These cases, they suggested, provide evidence of an interface between implicit and explicit knowledge.

Based on these findings, the authors proposed that all three interface positions can coexist and that the presence or absence of an interface depends on the difficulty of the structure, and consequently that certain structures are more likely to exhibit a permanent lack of interface. For example, structures that are sometimes associated with high accuracy on tests of explicit knowledge, but for which learners consistently make errors on tasks assessing implicit knowledge, such as inflectional morphemes or articles, may be more prone to permanent lack of interface.

In their view, the imbalance between implicit and explicit knowledge, as observed in their examples, reflects a lack of interface. However, because this imbalance varies across different areas of L2 acquisition, Han and Finneran argue for the simultaneous existence of multiple interfaces with varying degrees of interaction. This perspective offers a more nuanced explanation for the inconsistencies in empirical findings that traditional interface positions have struggled to address.

2.1.3. Measures of implicit and explicit knowledge

One of the primary goals in SLA research is to define and describe L2 linguistic knowledge and to explain how this knowledge develops over time, while considering other factors that may mediate or moderate that process (R. Ellis, 1994). As discussed in the preceding paragraphs, SLA researchers generally take the dual-system view, which distinguishes between implicit and explicit knowledge. In light of this, R. Ellis (2005) proposed a framework comprising seven criteria by which researchers can classify tests as targeting either implicit or explicit knowledge. These criteria are outlined in Table 1.

Table 1. Operationalizing the constructs of L2 implicit and explicit knowledge (taken from R. Ellis, 2005)

Criterion	Implicit knowledge	Explicit knowledge
Degree of awareness	Response according to feel	Response using rules

Time available	Time pressure	No time pressure
Focus of attention	Primary focus on meaning	Primary focus on form
Systematicity	Consistent responses	Variable responses
Certainty	High degree of certainty in responses	Low degree of certainty in responses
Metalinguistic knowledge	Metalinguistic knowledge not required	Metalinguistic knowledge encouraged
Learnability	Early learning favoured	Late, form-focused instruction favoured

As illustrated in Table 1, tests targeting implicit knowledge require learners to respond according to feel, with a high degree of certainty in their answers. Their responses should demonstrate consistency, and time pressure is applied to ensure that the task focuses on meaning rather than form. The targeted language features are typically those introduced early in the learning process, and learners are not expected to have metalinguistic knowledge (metalanguage or a language consisting of technical and semitechnical terminology used to describe another language, R. Ellis, 2009a) of the structures being tested. In contrast, tests of explicit knowledge require learners to respond based on rules they have learned, with potentially lower certainty in their answers. These tests often yield greater variability in scores and do not involve time pressure. The focus is on linguistic form, and the targeted structures are usually introduced later in the learning process. Additionally, these tasks encourage the use of metalinguistic knowledge. However, contrary to these hypotheses, subsequent studies have repeatedly demonstrated that systematicity in implicit tasks is similar to or even greater then in explicit tasks (R. Ellis, 2009b; Godfroid & Kim, 2021), while certainty has been found to be significantly and positively related to higher scores on tasks measuring explicit, rather than implicit, knowledge (R. Ellis, 2009b; Maie & DeKeyser, 2020).

Using this theoretical framework, R. Ellis (2005) conducted a psychometric study that included tests hypothesized to measure implicit knowledge – such as oral narrative, elicited imitation, and timed grammaticality judgment tests – and those hypothesized to measures explicit knowledge, such as untimed written grammaticality judgment tests and a metalinguistic knowledge test. The metalinguistic knowledge test was an untimed multiple-choice test where participants were presented with ungrammatical sentences and asked to select the rule that best explained each error from four provided options. This was followed by a second phase, during which participants read a short story and were required to identify specific grammatical features in the text, as well as name grammatical parts in a set of sentences. The untimed grammaticality judgment test asked participants to determine whether a sentence was grammatical or ungrammatical, indicate their certainty on a scale from 0% to 100% (as proposed by Sorace, 1996), and report whether their judgment was based on a rule or intuition. Both tests, as evident, lacked time pressure, directed participants' attention to form, and encouraged the application of metalinguistic rules.

A timed grammaticality judgment test was used to assess implicit knowledge. This version of the test did not require participants to indicate their degree of certainty or source attribution, in order to minimize focus on form. In the oral narrative test, participants were first instructed to read a story designed to elicit the use of target structures such as the regular past tense, modal verbs, and the indefinite article. They were then asked to retell the story orally within three minutes. Finally, an imitation test was employed, in which participants were presented with belief statements containing both grammatical and ungrammatical target structures. To direct attention to meaning, participants were asked to indicate whether they agreed with each statement after hearing it. They were then required to repeat the sentence in correct English. The tests were administered in order, progressing from more implicit to more explicit.

All tests demonstrated a reliability score above .8, which was considered satisfactory. Correlation analysis revealed significant interrelationships among all the measures, and an exploratory factor analysis identified two factors with eigenvalues greater than .8, labelled “implicit” and “explicit” knowledge. The imitation test, oral narrative test, and timed grammaticality judgment test loaded onto the implicit factor, together accounting for 58% of the variance. Meanwhile, the untimed grammaticality judgment test and the metalinguistic knowledge test loaded onto the explicit factor, explaining 16% of the total variance. Together, the two factors accounted for 74% of the total variance.

Godfroid and Kim (2021) explored a battery of nine tests as potential measures of implicit knowledge. They used the predictive power of various implicit aptitude measures to validate these nine tests. The logic behind their approach was that when performance on a linguistic task shares variance with an aptitude measure – meaning one can predict the other – a common cognitive process can be inferred to guide both tasks. Thus, a significant predictive relationship between a measure of implicit aptitude and a linguistic test would confirm the test as a valid measure of implicit knowledge. In their study, implicit aptitude was operationalized as implicit statistical learning, which was measured using the alternating serial reaction time (ASRT) task (Howard & Howard, 1997). The results showed that timed, accuracy-based tests like elicited imitation and timed grammaticality judgment tests were valid measures of implicit knowledge, while reaction-time tasks like self-paced reading and word monitoring were not.

Self-paced reading task requires participants to read grammatical and ungrammatical sentences in a word-by-word fashion. Participants progress through the sentences by clicking a button, and their reaction times are recorded. The final score is calculated as the difference in reaction times between grammatical and ungrammatical sentences, which is believed to reflect sensitivity to grammatical errors (Jiang, 2004, 2007) and tap into one’s implicit knowledge at a processing level. In a word monitoring task, participants are presented with a target word and have to respond to hearing this word in a sentence by pressing a button. The task includes both grammatical and ungrammatical sentences and the score is computed as the difference in reaction times between the two types of sentences. Similar to the self-paced reading task, the delay in response time for ungrammatical sentences is thought to reflect implicit knowledge. The findings of the Godfroid and Kim (2021) thus suggest that such reaction-time-based measures are not reliable indicators of implicit knowledge, while time-pressured, form-focused tasks are more appropriate measures, corroborating earlier research (R. Ellis, 2005).

However, findings from other studies challenge these conclusions. In a study by Vafaee et al. (2017), both the self-paced reading and word monitoring tasks loaded together on a factor distinct from the one on which the timed and untimed grammaticality judgment tests, as well as a test of metalinguistic knowledge, were grouped. This was interpreted as evidence that reaction-time measures, which tap into the processing phase of implicit knowledge and direct participant' attention to meaning, are more valid as measures of implicit knowledge than behavioural measures such as (un)timed grammaticality judgment tests.

Further, findings from Suzuki and DeKeyser (2015) challenge the status of elicited imitation as a measure of implicit knowledge. Their study found no predictive relationship between implicit aptitude, operationalized as implicit sequence learning ability via the serial reaction time (SRT) task (Kaufman et al., 2010), and elicited imitation. Instead, a significant relationship was observed between elicited imitation and explicit aptitude, measured as language analytic ability using the LLAMA F subtest of the LLAMA battery (Meara, 2005; Meara & Rogers, 2019). Based on this, the researchers concluded that elicited imitation likely gauges highly automatic explicit knowledge rather than implicit knowledge. According to Suzuki (2017), this type of knowledge is a form of conscious linguistic knowledge with varying levels of automatization. While this knowledge is functionally similar to implicit knowledge, it remains conscious, aligning with DeKeyser's (2015, 2020) Skill Acquisition Theory, which posits that the automatization of explicit knowledge occurs gradually over time.

In a subsequent study, Suzuki (2017) found that three timed, accuracy-based measures of L2 knowledge – auditory and written timed grammaticality judgment tests and a timed gap-fill test – loaded separately from three real-time comprehension tasks, such as self-paced reading, word monitoring, and the visual-world task. In the visual-world task, participants are presented with four images on a screen while they listen to a short story with their eye movements being tracked. The sentences are designed so that if participants are sensitive to the target feature of the task, they would spend more time looking at the target image compared to the competitor image after hearing a linguistic marker in the target condition. Thus, the separate loadings in Suzuki (2017) support the claim that time pressure alone is not sufficient for a measure to tap into implicit knowledge. In fact, since these measures likely tap into automatized explicit knowledge, Suzuki and DeKeyser (2015) argue that awareness should be the criterion used to differentiate between measures of implicit and automatized explicit knowledge. They suggest that time-pressured tasks do not necessarily limit access to explicit knowledge enough to ensure that only implicit knowledge is used (DeKeyser, 2003; Suzuki & DeKeyser, 2015; Vafaee et al., 2017). Instead, fine-grained online comprehension measures, such as self-paced reading and word-monitoring tasks, which capture linguistic knowledge at the processing level, should be employed.

In a self-paced reading task, half of the sentences are typically grammatically correct, and half incorrect. The score is calculated as the difference in reaction time between grammatical and ungrammatical sentences. This task provides detailed information about processing and can detect where and when additional processing is needed (Mackey & Gass, 2022). This is crucial, as additional processing or difficulty indicates sensitivity to grammaticality, or, in other words, "knowledge of and/or sensitivity to linguistic phenomena"

(Marsden et al., 2018, p. 862). In a word-monitoring task, participants listen to a sentence waiting to hear a target word that is displayed to them beforehand. They respond by pressing a button when they hear the word in the sentence. As in the self-paced reading task, half of the sentences are grammatically correct and half are incorrect. The score is determined by the difference in reaction time between the two conditions. In both tasks, a higher score reflects greater hesitation or stronger implicit knowledge of the targeted feature.

2.2. Language aptitude

As noted earlier, the validation of linguistic tests as measures of implicit or explicit knowledge is often based on their relationship to the cognitive aptitude believed to underlie the acquisition of that knowledge. However, there is also a lack of consensus in research on what exactly constitutes language aptitude and what its key components are. Language learning aptitude is often described as a broad, catch-all term that refers to various cognitive and perceptual abilities contributing to successful language learning (Granena, 2020). Since the advent of aptitude research during the Cold War era, scholars in SLA have been particularly interested in the predictive power language aptitude. The ability to measure language aptitude was seen as a way to predict an individual's long-term achievement and to offer insight into why some learners progress rapidly and achieve advanced levels of L2 proficiency, while others struggle persistently. Since Carroll and Sapon's (1959) landmark study, our understanding of language aptitude has evolved considerably, with researchers shifting their focus from its predictive role to its explanatory role in L2 learning. The following sections introduce the concept of language aptitude and review the history of empirical research. Prominent theoretical models and corresponding measurement batteries are reviewed. Lastly, contemporary views on language aptitude – specifically as it relates to implicit and explicit learning – are explored, along with an examination of the mediating and moderating roles of factors like proficiency and structural difficulty, which have often been overlooked in previous research.

2.2.1. The concept of language aptitude

Many researchers view language aptitude as mainly a cognitive variable with specific capacity for learning a second language, generally believed to involve several distinct abilities, such as auditory, linguistic, and memory ability (Skehan, 1998). Others stress the importance of affective and conative factors for L2 learning success (Ackerman, 2003; Dörnyei, 2005; MacIntyre, 2002; Snow, 1987) with such view aligning with Snow's (1994) notion of the "cognitive-affective-conative triangle," suggesting that language aptitude is shaped by this interplay of factors. A meta-analysis examining correlational research on explicit aptitude and L2 achievement from the 1960s to the 2010s found that overall proficiency was moderately correlated with cognitive aspect of aptitude ($r = 0.5$). In contrast, working memory ($r = 0.27$), motivation ($r = 0.37$), and anxiety ($r = -0.36$) were only weakly correlated with proficiency (Li, 2022). This reaffirms the importance of cognitive aptitude in language success, beyond conative and affective factors, thus, while the important role of affective and conative factors is acknowledged, the current research will focus on the cognitive side of language aptitude and the term "aptitude" will consistently throughout the thesis refer to the cognitive and perceptual aspects only.

2.2.2. Early research

In the 1950s and 1960s, John Carroll conducted pioneering research on language aptitude, much of which was funded by the U.S. government with the goal of establishing measures to predict the pace of language acquisition. These measures were intended to be utilized by government organizations for selection purposes in language programs (Granena, 2020). Carroll's work led to the development of the Modern Language Aptitude Test (MLAT; Carroll

& Sapon, 1959), which remains the most well-known and widely used language aptitude measure to date. The MLAT served as a benchmark for subsequent language aptitude assessments, such as the Language Aptitude Battery (PLAB; Pimsleur, 1966) and Defense Language Aptitude Battery (DLAB; Petersen & Al-Haik, 1976). The shared objective of these tests was to maximize predictive validity, allowing them to forecast future language learning outcomes (Granena, 2020).

Carroll adopted a bottom-up approach to developing the final MLAT, which consists of five subtests: number learning, phonetic script, spelling clues/hidden words, words in sentences, and paired associates. The number learning subtest measures participants' memory and auditory acuity, while the phonetic script subtest evaluates the ability to associate sounds with symbols. The words in sentences subtest gauges participants' grammatical sensitivity without using grammatical terminology, and the paired associates subtest measures rote memory (Carroll, 1962). Since no established theory of aptitude existed prior to Carroll's research, he proposed his aptitude model based on empirical analysis. His final model identified four key components or quantifiable skills: phonetic coding ability (tested via Phonetic Script), associative memory (measured by Paired Associates), grammatical sensitivity (assessed through Words in Sentences), and inductive language learning ability. While the MLAT examines three of these components, the inductive language learning ability was included in the initial battery of tests but not in the final set of five subtests. An overview of the proposed aptitude components can be seen in Table 2.

Table 2. Carroll's four-component model of aptitude (taken from Dörnyei & Skehan, 2003)

Aptitude components	Definitions
Phonetic coding ability	Capacity to code unfamiliar sounds so that it can be retained over a period of time and ultimately retrieved or recognized

Grammatical sensitivity	Capacity to identify the functions of words in sentences
Associative memory	Capacity to form associative links in memory between vocabulary items in L1 and L2
Inductive language learning ability	Capacity to extrapolate syntactic and morphological patterns from language and use this to create new sentences

From a practical standpoint, the MLAT proved to be an effective predictor of the rate of advancement in foreign language classrooms, with its predictive validity ranging from 0.25 to 0.83 (Granena, 2020). However, success in these studies was typically measured by course grades, which were largely based on quizzes and written exams. Given that the dominant teaching approach during that period was the audio-lingual method, it is unsurprising that cognitive abilities related to memory had the highest predictive validity, as four of the five MLAT subtests assess distinct memory components. Despite this, more recent research by Erlam (2005) demonstrated that the predictive role of language aptitude, as measured by The Sound Discrimination test from the PLAB and the Words in Sentences subtest from the MLAT, remains relevant under different learning conditions, including in communicative classroom environments. What is more, a meta-analysis of 34 studies conducted over 50 years (1963–2013) by Li (2015) confirmed that MLAT scores consistently and significantly predict L2 learning outcomes, with a positive correlation of r = 0.34. Therefore, Carroll's concept of language aptitude and the MLAT battery have shown enduring relevance, with Carroll's model continuing to be the most prominent and influential model of language aptitude to date (Skehan, 2002).

2.2.3. Recent developments

2.2.3.1. Intelligence perspective

Grigorenko and colleagues (2000) introduced a new interpretation of language aptitude, termed the Cognitive Ability for Novelty in Language Acquisition-Foreign (CANAL-F) theory, which is based on Sternberg's (1997, 2002) "successful intelligence" framework. In this model, intelligence is a central piece and it's viewed as consisting of three distinct components: analytical, creative, and practical (Grigorenko et al., 2000). The core idea is that successful language learners are those who can effectively handle novelty and ambiguity in the L2 learning process. To assess this, the authors developed the CANAL-F exam, which measures learners' ability to recall and infer novel language materials in an artificial language called "Ursulu" under immediate and delayed conditions (Dörnyei & Skehan, 2003). Five distinct information acquisition processes – selective encoding, accidental encoding, selective comparison, selective transfer, and selective combination – are operationalized at the lexical, morphological, semantic, and syntactic levels (Grigorenko et al., 2000). The test comprises five components. The first focuses on learning the meanings of neologisms from context, where participants are presented with 24 brief paragraphs orally and visually and must choose the correct English meaning for each unfamiliar neologism. The second section assesses general text comprehension. In the third component, participants engage in paired-associate learning, where they must memorize 60 word-pairs in English and Ursulu. The fourth component presents Ursulu sentences with English translations, and participants are asked to choose the best translations for new phrases in both directions. The final section involves a 12-item quiz to assess participants' overall understanding of Ursulu.

The authors argue that this test offers several advantages over traditional aptitude tests, including (a) its cognitive and theoretical grounding, (b) its contextualized and dynamic nature, (c) its multifunctional approach, and (d) its adaptability. Additionally, the CANAL-F emphasizes both working memory and long-term memory, which are assessed through immediate and delayed recall tasks. However, despite its innovative design, validation studies showed that the CANAL-F did not significantly outperform the MLAT in terms of predictive validity (Sternberg & Grigorenko, 2002). More importantly for researchers, the test is no longer available.

2.2.3.2. Linguistic Coding Differences Hypothesis

Sparks and colleagues (1991, 2001) proposed the Linguistic Coding Differences Hypothesis (LCDH) as an alternative model of language aptitude. The LCDH is grounded in the idea that L1 literacy skills are critical predictors of L2 learning success. For example, difficulties in L1 phonology are likely to negatively impact a learner's ability to acquire an L2 (Sparks & Ganschow, 2001). A factor-analytic study by Sparks et al. (2011) revealed that 76% of the variance in ultimate L2 oral and written proficiency can be accounted for by four key components: (1) L1 and L2 phonology/orthography skills, (2) L1 and L2 language analytical skills, (3) IQ/memory skills, and (4) self-perceptions of L2 motivation and anxiety. The authors emphasize the importance of examining both the similarities and differences between learners' L1 and L2, particularly in relation to potential negative transfer effects. In this sense, the LCDH complements Carroll's four-factor model of language aptitude by introducing the additional component of "L1 and L2 phonology/orthographic decoding skill" (Wen et al., 2017).

2.2.3.3. Information processing perspective

Skehan (2002, 2012, 2016) proposed a model of language aptitude, often referred to as the staged model, which builds on developments from SLA research. His model is based on Carroll's (1981, 1993) four-component framework, but Skehan merged grammatical sensitivity and inductive language learning ability into a broader construct called language-analytic ability. He further suggested that different components of aptitude should be aligned with distinct stages of SLA development and their corresponding cognitive processes. The final model consists of nine stages, as shown in Table 3.

Table 3. Skehan's staged aptitude model

SLA stages	L2 cognitive processes	Aptitude constructs
Language input	Input processing	Attentional control
Central processing	(segmentation)	Working memory
Language output	Noticing	Phonetic coding ability
		Working memory
	Pattern recognition	Phonetic coding ability
		Working memory
		Language analysis ability
	Complexification	Language analysis ability
		Working memory
	Handling feedback	Language analysis ability
		Working memory
	Error avoidance	Working memory
		Retrieval Memory
	Automatization	Retrieval Memory
	Creating a repertoire	Retrieval Memory
		Chunking
	Lexicalization	Chunking

Note: Attentional control, working memory, retrieval memory, and chunking are new components in comparison with Carroll’s model.

Skehan argued that various components of language aptitude become relevant at different stages of development. In the latest and final iteration of his model (Skehan, 2016), phonetic coding ability is emphasized during stages 2 and 3, while language-analytic ability is central during stages 3 to 5. Notably, while working memory was initially considered important only during the early stages (Skehan, 2002), it is now viewed as relevant across stages 1 to 6 in the 2016 model. Additionally, retrieval memory plays a key role during stages 6 to 8. This staged approach implies that the role of different aptitude abilities varies depending on the learner's developmental stage, with specific components becoming more or less important over time.

Skehan (2016) also proposed that the nine stages can be broadly divided into two phases: the first half (stages 1-5) focuses primarily on knowledge acquisition, while the second half (stages 6-9) is concerned with the control of knowledge in performance, as well as the ability to sound develop pronunciation (Wen et al., 2017). This approach to language aptitude shifts the focus from merely predicting language learning outcomes to providing a deeper explanation of the underlying processes, resulting in significant implications for the field. Skehan's model prioritizes explanatory power over predictive accuracy, offering a more nuanced understanding of how aptitude influences language learning across different stages.

2.2.3.4. Aptitude-treatment interaction

Robinson (2005, 2007, 2012) developed the Aptitude Complexes/Ability Differential framework, sometimes referred to as the ATI model, based on Snow's interactionist approach (1994). This framework identifies various aptitude complexes, or combinations of cognitive abilities, that are differentially related to processing under varying conditions of instructional exposure to L2 input (Robinson, 2007). The model is built upon two key hypotheses: (1) the Aptitude Complexes Hypothesis and (2) the Ability Differentiation Hypothesis.

The Aptitude Complexes Hypothesis posits that basic cognitive abilities combine to form higher-order aptitude complexes, which are drawn upon during specific tasks. This suggests that L2 learners exhibit variations in their basic cognitive abilities, resulting in differentiated profiles within these aptitude complexes. Examples of basic cognitive abilities include processing speed, pattern recognition, semantic priming, and phonological working memory capacity. Higher-order aptitude complexes might include constructs such as noticing the gap, memory for contingent speech, and metalinguistic rule rehearsal (Robinson, 2005). The Ability Differentiation Hypothesis asserts that learners display individual differences in cognitive abilities, which lead to variations in their corresponding aptitude complexes and, ultimately, differentiated aptitude profiles. Together, these hypotheses form the basis of Robinson's framework, often represented schematically by a wheel-shaped diagram with multiple layers of embedded circles (Figure 1, taken from Robinson, 2005).

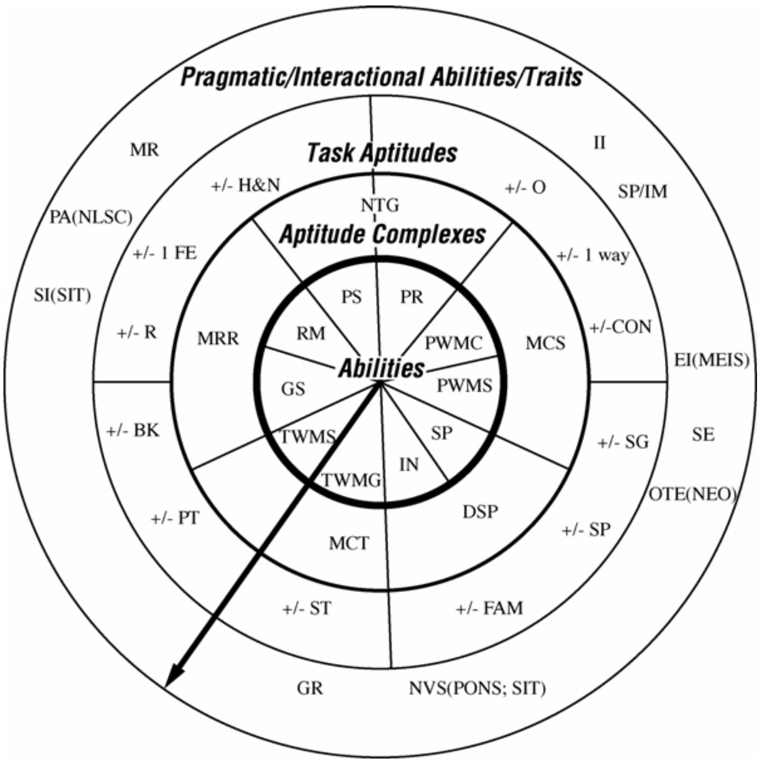

Figure 1. ATI model of aptitude (taken from Robinson, 2005)

The innermost circle of Robinson's framework contains core abilities or aptitudes, such as grammatical sensitivity and processing speed, while the second circle focuses on aptitude complexes. These complexes include an aptitude for focusing on form, an aptitude for incidental learning via oral content, incidental learning via written content, and an aptitude for explicit rule learning. The outer two circles (third and fourth) address instructional factors and contexts, as well as broader communicative influences and capacities (Robinson, 2005). Abilities in the inner two circles are involved in “initial input-based learning,” the third circle relates to “output practice and complex task performance,” and the outermost circle concerns the “transfer of task performance to real-world interactive settings” (Robinson, 2005, p. 52).

The core of Robinson’s model lies in the two inner circles, which focus on fundamental abilities and aptitude complexes. Abilities represent essential aptitudes comparable to those in earlier aptitude models, while the aptitude complexes circle is a novel feature. Specifically, the model identifies four key aptitude complexes: an aptitude for focus on form, an aptitude for incidental learning via oral content, an aptitude for incidental learning via written content, and an aptitude for explicit rule learning.

Each aptitude complex is influenced by two ability factors that can have high or low values, resulting in four possible combinations for each complex (Robinson, 2007). For example, the aptitude for explicit rule learning is shaped by (1) memory for contingent text and (2) metalinguistic rule rehearsal, with these factors varying between high and low levels (e.g., high memory for contingent text and high metalinguistic rule rehearsal, or high memory for contingent text and low metalinguistic rule rehearsal). This indicates that Robinson views aptitude as a dynamic phenomenon, dependent on the context of learning and instructional conditions. In other words, the model suggests that various aptitude components may interact differently depending on the specific learning context.

Robinson’s ATI model shares similarities with Skehan’s developmental model in that both seek to explain observable individual differences in L2 learners’ final learning outcomes. However, Robinson’s model differs in its finer-grained representation of complex cognitive abilities, incorporating several levels of analysis. Basic cognitive abilities form the deeply embedded core, while aptitude complexes occupy a less “core” level, and task aptitudes and actual performance represent the least core, most real-world level. Additionally, the model emphasizes the dynamic nature of language aptitude, as aptitude complexes depend on the interaction of their core components and the specific characteristics of the learning task. Essentially, the ATI model accounts for aptitude-treatment interaction (ATI), which occurs across all stages of the SLA process (Wen et al., 2017).

2.2.3.5. Cognitive science perspective

Many L2 learners begin acquiring a second language as adults, and research has consistently shown that learners that started learning an L2 as adults can be distinguished from those who began language acquisition early in life (Abrahamsson & Hyltenstam, 2009; DeKeyser & Larson-Hall, 2005; Long, 2005). Nevertheless, some L2 learners attain high proficiency despite a late start. This observation led to the development of the High-level Language Aptitude Battery (Hi-LAB), designed to measure language aptitude in talented learners who reach advanced levels of proficiency.

Empirical evidence from Linck et al.'s (2013) study which relied on the Hi-LAB test, suggests that the factors contributing to high-level proficiency may differ from those influencing earlier stages of learning. The study proposed that high-level aptitude should be viewed as "a composite of domain-general cognitive abilities and specific perceptual abilities" (Linck et al., 2013, p. 535). This factor-based model aligns with contemporary cognitive science, with the authors identifying working memory as a key contributor to language aptitude. The constructs and measures employed in the Hi-LAB test, as outlined in Table 4, support this view.

Table 4. Hi-LAB constructs and measures (taken from Linck et al. 2013)

Constructs	Measures
Working memory	
Executive functioning	

Updating	Running memory span
Inhibitory control	Antisaccade
Task switching	Task-switching numbers
Phonological short-term memory	Letter span
	Non-word span
Associative memory	Paired associates
Long-term memory retrieval (priming)	ALTM synonym
Implicit learning	Serial reaction time
Processing speed	Serial reaction time
Auditory perceptual acuity	Phonemic discrimination: Hindi, English
	Phonemic categorization: Russian

Note: ALTM = Available long-term memory

As shown in Table 4, working memory is assessed using five separate measures, each probing different components: executive working memory (EWM) and phonological working memory (PWM).

An interesting finding from the study was that the most effective sub-tests were the paired associates test, the letter span test, and the serial reaction time test, while the non-word span test was only marginally significant (Linck et al., 2013). In other words, measures of central executive memory, processing speed, and auditory abilities were equally relevant for both successful and highly successful learners. Moreover, Linck et al.'s findings support the importance of working memory as outlined in Skehan's theory, as the more effective measures in the study were those of working memory, consistent with Skehan's view that retrieval memory becomes significant during later stages of SLA.

In summary, the notion and conceptualization of language aptitude have undergone numerous revisions since its inception. While the core components of the Carrollian model remain prominent in more recent frameworks, various alternative models and modifications to the original concept have been proposed over the years. Both Skehan's (2002, 2012, 2016) and Robinson's (2005, 2007, 2012) models of aptitude represent efforts to position aptitude within the broader context of language learning, emphasizing its role as part of a developmental process. Robinson's model, in particular, highlights the influence of external factors that can shape the effects of aptitude, while its multilayered structure suggests that combinations of more fundamental cognitive abilities have a far-reaching impact on the higher-order abilities critical for language learning. Although these models are theoretically compelling, they lack accompanying tests to operationalize their constructs. Similarly, attempts to define the role of intelligence and the effects of L1 on language aptitude for L2 learning have not resulted in accessible or widely available testing methods. While the recent Hi-LAB test shows promise, it too is not publicly available for researchers. Finally, a notable limitation of the tests reviewed is their exclusive focus on L1 speakers of English, which runs counter to ongoing efforts to broaden the scope of SLA research to other languages.

2.2.3.6. Modernizing aptitude assessments: The LLAMA suite

The LLAMA test battery of language aptitude (Meara, 2005; Meara & Rogers, 2019) addresses the above issues quite successfully. It is strongly influenced by Carroll's notion of aptitude, which comprises four components. Unlike the MLAT, the LLAMA subtests are independent of a specific L1, making them usable across different linguistic contexts. Additionally, the LLAMA is freely available and easy to administer, which has contributed to its widespread use among researchers. The LLAMA battery consists of four subtests: LLAMA B, LLAMA D, LLAMA E, and LLAMA F.

LLAMA B serves as the vocabulary learning module and is a basic associative memory task designed to measure how quickly one can acquire new vocabulary. It is similar to the paired associates subtest of the MLAT but has the advantage of using unfamiliar graphical objects and not using words from any real language (Rogers et al., 2017).

LLAMA D is a unique subtest not present in the MLAT. It assesses the capacity to recognize segments of spoken language after brief exposure. According to its developers, this ability is crucial for language acquisition, as learners who can identify repeated sound sequences are better equipped to detect subtle differences in speech, facilitating the identification of specific words and morphological variants (Rogers et al., 2017). This subtest is said to measure auditory pattern recognition ability.

LLAMA E is an adaptation of the MLAT's sound-symbol correspondence task. It evaluates an individual's ability to create associations between sounds and symbols. This subtest is said to gauge phonetic coding ability.

LLAMA F is a grammatical inference test in which participants have to deduce the grammatical rules of an unfamiliar language. This task requires participants to use their inductive language-analytic ability to infer aspects of the language's grammar, such as word order, gender, number (singular, dual, plural), and prepositions.

Since 2005, despite its popularity, the LLAMA battery has been subjected to only a limited number of validation studies. While some studies offer promising evidence regarding its validity and reliability (Rogers et al., 2023; Rogers et al., 2016), others have identified certain limitations (Bokander & Bylund, 2020). Nevertheless, the creators of the LLAMA battery have addressed several of these limitations in the latest version (version 3) and have expressed their commitment to ongoing updates, with future improvements likely (Rogers et al., 2023). Additionally, the increasing number of studies employing the battery in recent years reflects its growing popularity and accessibility. This trend facilitates comparisons of results between studies and suggests further increase in use in the future.

2.2.4. Implicit and explicit language aptitude

The distinction between aptitude for implicit and explicit learning is a relatively recent development (Granena, 2016, 2020; Li, 2022; Li & DeKeyser, 2021; Linck et al., 2013). Traditionally, following the Carrollian view, language aptitude consists of components such as associative memory, the ability to memorize word-meaning associations; phonetic coding, the ability to recognize sounds and learn sound-symbol correspondences; and language-analytic ability, the capacity to identify grammatical functions of words in sentences and extrapolate linguistic regularities (Li, 2022). These abilities have been shown to correlate with the effects of explicit instruction, where learners engage in intentional and conscious learning (Li, 2015). For this reason, these are considered the foundational abilities underlying aptitude for explicit learning (hereafter, explicit aptitude).

Aptitude for implicit learning (hereafter, implicit aptitude) refers to cognitive abilities that facilitate implicit L2 processing and learning in the absence of conscious awareness (Granena, 2020). The advancements related to implicit aptitude are closely linked to developments in cognitive psychology concerning implicit learning. Historically, researchers believed that implicit learning was evolutionarily older, leading to the assumption that there would be little individual variation (Reber, 1967, 1992; Stanovich, 2009), with any variation dismissed as noise or unexplainable (Kaufman et al., 2010; see Section 2.1.1.1). However, recent evidence suggests systematic variation among individuals (Kaufman et al., 2010; Misyak & Christiansen, 2012; Misyak et al., 2010), indicating that abilities related to implicit learning may have predictive power in language acquisition, as many SLA theories view language learning as a largely implicit process (N. Ellis, 2005; Ellis, 2015; Krashen, 1981, 1985; Long, 1996, 2015).

The distinction between implicit and explicit aptitude, along with their respective roles in language learning, aligns well with prominent SLA theories (see Section 2.1.1.3). From a usage-based perspective, implicit learning – and by extension, implicit aptitude – plays a central role given that the majority of language learning happens implicitly. Similarly, the Interaction Hypothesis posits that implicit learning, and hence implicit aptitude, is the default pathway for L2 acquisition, particularly in children. However, as implicit aptitude weakens in adults, explicit learning (and explicit aptitude) helps to compensate for this. In contrast, Skill Acquisition Theory suggests that implicit aptitude becomes crucial in later stages of L2 development, specifically during procedural and automatic phases, while explicit aptitude is more relevant at early stages of learning. Finally, the Declarative-Procedural Model links implicit aptitude to procedural memory which is responsible for learning grammar and parts of language that follow patterns, while explicit aptitude is related to declarative memory which is responsible for learning idiosyncratic information and arbitrary associations.

Like explicit aptitude, which is understood to be multi-componential both theoretically and empirically, implicit aptitude is also conceptualized as having multiple components (Granena, 2020; Li & DeKeyser, 2021). The strongest evidence comes in the form of the lack of convergent validity among measures of implicit learning, likely reflecting the existence of different pathways through which implicit learning occurs (DeKeyser & Li, 2021).

2.2.5. Predictive validity of implicit and explicit aptitude

The strongest evidence supporting the predictive role of explicit aptitude comes from Li's (2016) meta-analysis. Drawing on data from more than 66 studies, the analysis demonstrated that the phonetic coding component of explicit aptitude is the strongest predictor of overall proficiency and vocabulary, compared to other components. Additionally, the language-analytic component was found to be the most significant predictor of grammar learning and L2 comprehension. Contrary to common assumptions and despite the prominent role in various aptitude models (Robinson, 2005, 2007, 2012; Skehan, 2002, 2012, 2016; Wen, 2016; Wen,

2019) and empirical studies (Linck et al., 2013; Linck & Weiss, 2011; Morgan-Short et al., 2014), associative memory showed only a weak relationship with general proficiency and specific aspects of L2 learning, suggesting a more limited role in L2 acquisition.

In contrast, implicit aptitude is a more recent concept, and studies examining its predictive power are still relatively scarce. However, Kaufman et al. (2010)showed that implicit sequence learning ability predicted students' performance in foreign language classes. Granena (2019) reported that LLAMA D could be considered a measure of implicit learning, as it loaded together with semantic priming scores, which were predictive of speech fluency. In this study, which included intermediate learners of L2 Spanish, the SRT (serial reaction Time) score did not significantly predict speaking proficiency. Linck et al. (2013), on the other hand, using the Hi-LAB battery provided additional evidence for the predictive role of implicit aptitude, operationalized as implicit sequence learning ability via the SRT, in highly advanced L2 learners of Spanish. Additionally, phonological short-term memory and associative memory were also significant predictors in this group. Saito et al. (2019), on the other hand, found that associative memory and phonetic coding ability – both components of explicit aptitude – significantly predicted learners' fluency during their first semester, while auditory pattern recognition ability, considered a component of implicit aptitude, was predictive of pronunciation scores in the second semester.

Both implicit and explicit aptitude have been shown to play predictive and explanatory roles in L2 learning. Although discrepancies exist between the measures used to assess them, evidence suggests that the roles of explicit and implicit aptitude may vary depending on factors such as the classroom setting, type of instruction, and learners' overall L2 proficiency.

2.2.6. L2 proficiency and language aptitude

Li's meta-analyses (2015, 2016) reveal that explicit aptitude is more strongly associated with L2 performance in high-school learners compared to university students. Given that high-school learners are likely less proficient, Li speculates that explicit aptitude may be more relevant at beginner levels of learning. This suggestion aligns with Carroll's (1990) perspective on language aptitude, which posits that explicit aptitude is primarily concerned with the rate of learning a language from the very beginning. However, this conclusion is not without its limitations. First, none of the studies included in the meta-analyses explicitly operationalized proficiency as a variable, rendering this claim speculative. Second, the studies included in the analysis varied considerably in terms of learning settings, outcome measures, and learners' proficiency levels, making it difficult to draw definitive conclusions. Nevertheless, a few studies conducted after the cut-off point for the meta-analysis (2013) did incorporate proficiency into their designs, either by examining the predictive role of aptitude at specific proficiency levels or by investigating the importance of aptitude factors across multiple proficiency levels.

Linck et al. (2013), for instance, used the Hi-LAB battery to investigate the role of cognitive and perceptual abilities at advanced proficiency levels. The Hi-LAB battery comprises 11 computer-delivered tests that tap into nine constructs, focusing primarily on working memory, with eight of the 11 measures assessing various aspects of working memory. The remaining measures assess implicit aptitude and phonemic discrimination. Unlike other aptitude measures such as the MLAT, DLAB, PLAB, and LLAMA, which assess traditional explicit aptitude, Hi-LAB does not assess phonetic coding ability or grammatical sensitivity.

Participants were categorized into mixed-attainment and high-attainment groups based on their scores on the Defense Language Proficiency Test (DLPT) and the demands of their occupations. The DLPT is a standardized proficiency test that measures listening and reading comprehension through multiple-choice questions administered in the participants' L1. Participants classified as high-attainment held two or more job assignments requiring a proficiency level of 4 or higher on the Inter-agency Language Roundtable scale. This scale consists of six levels of proficiency, which describe individuals' ability to function in a work environment using an L2.

Logistic regression analyses revealed that implicit aptitude (measured by the SRT), associative memory (assessed by Paired Associates), and phonological working memory (evaluated through Letter Span) were positively correlated with high language attainment. Interestingly, there was a negative correlation between the switching component of executive working memory (measured by Task Switching Numbers) and high attainment in listening. The authors attributed this finding to increased switching between the L2 and L1, leading to greater reliance on L1, which may hinder high L2 attainment. Overall, these findings suggest that implicit language aptitude and working memory play significant roles in language success at advanced proficiency levels.

In another study, Artieda and Muñoz (2016) used the LLAMA battery (Meara, 2005; Meara & Rogers, 2019) to examine the role of cognitive aptitudes at beginner and intermediate proficiency levels. Beginner learners were enrolled in the first year at a government-owned language school (CEFR level A1), while intermediate learners were in their fourth year (CEFR levels B1-B2). The age range of participants in both groups was wide, spanning from 16 to 62. On the global aptitude score, where scores from different subtests were combined, the effect of aptitude was the same at both proficiency levels, with a medium effect size. When examining each subtest individually, auditory pattern recognition (LLAMA D) had a significant predictive role at the beginner level only, while language-analytic ability (LLAMA F) was relevant at both proficiency levels. Associative memory (LLAMA B) and phonetic coding ability (LLAMA E) were only predictive at the intermediate level.

Morgan-Short et al. (2014) examined the predictive power of declarative and procedural memory at early and late stages of acquisition. As previously noted, declarative and procedural memory are not equivalent to explicit and implicit aptitude; rather, explicit and implicit aptitude are broader concepts, with declarative and procedural memory serving those cognitive systems. The Paired Associates subtest of the MLAT (Carroll & Sapon, 1959) was used as a verbal measure of declarative memory, while the Continuous Visual Memory Task (Trahan & Larrabee, 1988) was employed as a nonverbal measure of declarative memory. The Tower of London task (Kaller et al., 2011; Kaller et al., 2012; Unterrainer et al., 2003) and the Weather Prediction Task (Foerde et al., 2006) were used to tap into procedural memory. Participants were trained on the artificial language Broncato2 (Morgan-Short et al., 2010; Morgan-Short et al., 2012) through a computer-based game designed to facilitate implicit learning of word order. Participants were assessed at session 3 (early stage) and session 7 (late stage). The key findings were that declarative memory uniquely predicted timed auditory grammaticality judgment scores at the early stage of acquisition, while procedural memory predicted scores at the later stage.

Despite the clear differences between these studies in their operationalization of aptitude and their measures of proficiency, a common pattern emerges. Explicit aptitude, along with the corresponding declarative memory, appears to be relevant and predictive of language attainment at beginner levels or during the early stages of acquisition. Abilities such as phonetic coding ability and language-analytic ability are relevant at early to intermediate levels, while abilities involved in implicit learning, supported by procedural memory, become more important at advanced proficiency levels.

2.2.7. Structure difficulty and language aptitude

In the context of SLA, learning difficulty can be categorized as either subjective or objective (DeKeyser, 2003; Housen & Simoens, 2016). Subjective difficulty arises from the interaction between language features and individual learner characteristics and capacities. Additionally, other learner-related factors, such as a learner's L1 and its correspondence to L2, overall L2 proficiency, and various conative and affective factors (e.g., motivation and anxiety), also contribute to subjective difficulty (Housen & Simoens, 2016). Consequently, the same grammatical features may be perceived as differently difficult by learners depending on their cognitive abilities, motivation, anxiety levels, or L1 background.

In contrast, objective or feature-related difficulty refers to the inherent difficulty of linguistic structures that are more cognitively demanding for all learners, regardless of their individual profiles (Housen & Simoens, 2016). DeKeyser (2005) proposed a set of characteristics believed to influence the relative learning difficulty of aspects of L2 morphosyntax: lack of transparency in form-meaning mapping due to communicative redundancy and optionality of form, and opacity or reliability of form-meaning mapping. Other factors included frequency in the input, salience, and regularity of the form-meaning relationship. R. Ellis (2006) proposed a similar set of characteristics, with the added distinction between implicit and explicit learning difficulty. In his view, characteristics that affect implicit learning difficulty are frequency, salience, regularity, and processability, and a those that affect explicit learning difficulty are namely systematicity, technicality, and conceptual complexity of the form-meaning mapping.

Drawing on these criteria, Roehr and Gánem-Gutiérrez (2009a) developed a taxonomy of variables that gives a detailed overview of features impacting implicit and explicit learning difficulty. These are as follows:

- Frequency (how frequently a structure appears in the input) with high frequency decreasing implicit learning difficulty,
- Perceptual salience (how easily a structure is perceived auditorily in the input) with high perceptual salience decreasing implicit learning difficulty,
- Communicative redundancy (how much a structure contributes to the communicative intent of a message) with low redundancy decreasing implicit learning difficulty,
- Opacity (whether a form maps onto single or multiple meanings/functions and vice versa), with low opacity decreasing implicit learning difficulty,
- Schematicity (whether a structure is schematic or specific) with high schematicity decreasing both implicit and explicit learning difficulty,
- Conceptual complexity (the number of elements needed for a metalinguistic description) with low complexity decreasing explicit learning difficulty,
- Technicality of metalanguage (whether the metalanguage used in descriptions is familiar or abstract) with low technicality decreasing explicit learning difficulty,
- Truth value (whether a metalinguistic description applies without exception) with high truth value decreasing explicit learning difficulty.

To complement this framework, Collins et al. (2009) propose a pedagogical perspective, suggesting that more complex pedagogical rules lead to greater difficulty compared to simpler rules. This perspective is particularly relevant in instructed contexts where explicit conditions include the provision of metalinguistic rules. Finally, DeKeyser (2003) posits that learning difficulty can be conceptualized as a ratio between the inherent difficulty of a linguistic structure (objective difficulty) and the learner's ability to cope with it (subjective difficulty).

2.2.7.1. Empirical studies examining the moderating role of structure difficulty

If language aptitude influences how easy or difficult a linguistic feature is for a learner, L2 difficulty can be said to moderate the role of language aptitude. However, empirical research investigating the moderating role of L2 difficulty on the effects of language aptitude is extremely limited. In one such laboratory study, Robinson (2002) examined three grammatical structures of varying difficulty – locative, ergative, and incorporation in Samoan. Ergative was considered the most difficult since ergative marking is an unfamiliar concept in accusative languages like Japanese, the participants' L1. Noun incorporation, which is allowed is Samoan, refers to incorporation of nouns that function as direct objects directly into verbs (e.g. ave-taavale → drove-car). This was considered somewhat simpler but still difficult, as it does not exist in Japanese (or English). Locative, the easiest feature, was considered more familiar due to the similarity of locative markers and word order in Japanese. The study examined the roles of language aptitude and working memory, alongside different instructional types – implicit, explicit, and incidental. Language aptitude was assessed using the Language Aptitude Battery for Japanese (Sasaki, 1996), which is based on the MLAT and PLAB, while working memory was measured using the reading span test (Osaka & Osaka, 1992). The results revealed that language-analytic ability was crucial for the two more difficult structures. For ergative, both language-analytic ability and working memory played significant roles, while for noun incorporation, associative memory also contributed. In contrast, for the easiest structure, locative, associative memory and phonetic coding ability were more relevant.

Similarly, Yalçın and Spada (2016) conducted a classroom-based study to investigate the role of explicit aptitude components in relation to the difficulty of grammatical features. Grammatical difficulty was defined in terms of structural complexity and salience in the input. Past progressive was selected as an easy feature due to its transparent form-meaning relationship and high frequency in the input. In contrast, the passive voice was deemed more difficult due to its grammatical complexity and lower frequency in the input. Participants received four hours of instruction on both features, and language aptitude was measured using the LLAMA battery (Meara, 2005; Meara & Rogers, 2019). A written grammaticality judgment test and an oral production task were used to assess explicit and implicit knowledge, respectively.

The findings indicated that the language-analytic component of explicit aptitude was particularly relevant for the more difficult and less frequent grammatical feature (passive voice) on the test assessing explicit knowledge (written GJT). This led the researchers to conclude that learners need to engage their language-analytic abilities when processing less frequent structures. In contrast, the associative memory component was deemed more relevant for the easier and more frequent feature (past progressive) on the test assessing implicit knowledge, suggesting that simpler and more frequent structures rely more heavily on memory abilities.

Although the available evidence is limited, these studies suggest that language aptitude components and working memory contribute differently depending on the difficulty of the L2 structure. For easier structures, learners tend to rely on associative memory and, occasionally, phonetic coding ability, while more difficult language features often require greater reliance on language-analytic ability and, frequently, working memory. These results support Robinson's dynamic view of language aptitude (2005, 2007, 2012), which conceptualises aptitude as context-dependent, proposing that different components of aptitude become more or less relevant depending on external factors, such as the difficulty and frequency of the target feature.

2.2.8. Measures of implicit aptitude

Measuring implicit aptitude presents significant challenges. The key issue is the ongoing debate about whether any task can truly be considered fully implicit, especially given that tasks measuring implicit learning consistently produce lower reliability scores compared to those assessing explicit learning. For instance, the serial reaction time task, one of the most commonly used measures of implicit learning, typically generates reliability indices between .40 and .50 (Granena, 2013b; Kaufman et al., 2010; Suzuki & DeKeyser, 2015, 2017) which is regarded as standard in the literature (Dienes, 1992; Reber et al., 1991). Granena (2020) further highlights that low reliability does not necessarily imply low validity, as reliability pertains to the scores generated by the measure rather than the measure itself. Supporting this, Kalra et al. (2019) report moderate test-retest reliability for SRT scores, suggesting some promise despite the lower split-half reliability indices. However, other measures are not as widely used, and thus their reliability and validity are yet to be thoroughly tested. The following paragraphs provide an overview of the most widely used measures for probing implicit aptitude within the context of SLA.

2.2.8.1. Serial Reaction Time task

The SRT task is currently the most widely used measure of implicit aptitude, specifically gauging implicit sequence learning ability (Granena, 2020). During the task, participants respond to continuous visual stimuli by pressing the appropriate response button as quickly and accurately as possible. Implicit learning is measured by comparing reaction times between a high-frequency sequence, which participants gradually respond to more quickly, and a random or alternate sequence that they have had less exposure to. The difference in reaction time reflects the degree of implicit learning that has occurred.

There are three main versions of the task: (1) deterministic, (2) alternating, and (3) probabilistic. In the alternating and probabilistic versions, transitions between sequences are governed by probabilities, whereas both deterministic and alternating versions use random sequences as baselines.

In the deterministic version (Nissen & Bullemer, 1987), participants are repeatedly presented with the same sequence of ten or twelve elements, with a random sequence introduced toward the end as a baseline. Faster reaction times in the last block for the repeating sequence compared to the random sequence indicate implicit learning.

The alternating version (Howard & Howard, 1997) presents stimuli in triplets, with random stimuli interspersed among predetermined stimuli that appear with higher frequency, reflecting higher probability. The score is calculated as the difference in reaction time between the random and predetermined stimuli.

In the probabilistic version (Kaufman et al., 2010), participants respond to a sequence that follows the training condition 85% of the time and the control condition 15% of the time. These sequences differ in the second-order conditional information they convey. To predict the next stimulus, the last two trials are needed, and there is an 85% probability that the next trial follows the training sequence, and 15% that it follows the control sequence. This version is thought to simulate the noisy conditions in which implicit learning usually occurs (Granena, 2020; Jiménez & Vázquez, 2005). The difference in reaction time between the two conditions is used to measure implicit learning, as participants are unaware of the existence of two sequences.

The SRT task is a visual, non-verbal measure that relies on participants' motor skills and assesses on-task learning (Christiansen, 2019). It has been widely employed in both validation and predictive studies (Godfroid & Kim, 2021; Granena, 2013a; Granena & Yilmaz, 2019; Hamrick, 2015; Iizuka & DeKeyser, 2023; Kalra et al., 2019; Kaufman et al., 2010; Linck et al., 2013; Roehr-Brackin et al., 2024; Roehr-Brackin et al., 2023; Suzuki & DeKeyser, 2015, 2017).

2.2.8.2. Artificial Grammar Learning Task

The Artificial Grammar Task (Reber, 1967) involves two phases: a learning phase and a testing phase. In the learning phase, participants are presented with a series of letter strings (e.g., TPPTXXVS) generated using a Markov process, where each subsequent letter is probabilistically dependent on the current letter. These letter strings are created from a limited set of five letters (P, S, T, V, X), and participants are instructed to memorize them. In the testing phase, participants are informed that an underlying grammatical system governed the letter strings, and they are then asked to judge whether a new set of unseen letter strings follows this "grammar." Afterward, participants are asked to verbalize any grammar rules they may have identified. Implicit learning is confirmed if participants' judgments of grammaticality are significantly above chance and if there is no correlation between their accuracy and their reportable knowledge of the grammar rules.

2.2.8.3. Visual Statistical Learning

This task (Frost et al., 2013) is designed to assess participants' ability to detect regularities in visual stimuli. In the task, participants are exposed to a series of shapes that follow adjacent contingencies, meaning that specific shapes consistently appear in a particular sequence (e.g., shape Y always follows shape X, and shape Z always follows shape Y). During the test phase, participants are presented with two-alternative forced-choice questions, where they must identify the original triplets that follow the learned contingencies from distractor sequences that violate these patterns. Suggested improvements for the task have included the addition of a confidence scale for each response and modifying the two-choice format to a "three correct and one correct missing" format for greater sensitivity.

2.2.8.4. Auditory statistical learning task

The Auditory Statistical Learning task (Siegelman et al., 2018) mirrors the structure of its visual counterpart, but with auditory stimuli. In this task, participants are exposed to an unsegmented stream of tri-syllabic nonsense words, presented as though they are listening to a monologue in an unfamiliar language. During the test phase, participants engage in a two-alternative forced-choice format, where they are required to identify which of the tri-syllabic words does not belong to the unfamiliar language. A notable limitation, as acknowledged by the authors, is that performance on the task can be influenced by participants' L1, potentially affecting their ability to discern the statistical patterns in the auditory stream.

2.2.8.5. Weather prediction task

This task (Knowlton et al., 1996; Knowlton et al., 1994) requires participants to predict either rain or sunshine based on geometric patterns that are probabilistically linked to each weather outcome. This task is a classic example of probabilistic category learning, where participants gradually learn to associate specific patterns with a higher likelihood of certain weather outcomes. Each trial provides feedback, and participants typically start by performing at chance level (around 50% accuracy) and finish closer to 70%, reflecting implicit learning without their ability to explicitly state the rules governing the probabilities. The task has been critiqued for allowing participants to develop explicit strategies, which some researchers have attempted to mitigate by incorporating secondary tasks, such as counting high tones, to reduce the likelihood of explicit strategy use (Morgan-Short et al., 2014).

2.2.8.6. Tower of London

The Tower of London task (Shallice 1982) presents participants with a board containing three sticks and several coloured balls. The objective is to rearrange the balls to match a target pattern displayed on another board by moving them between the sticks, with the goal of minimizing the number of moves. This task requires participants to plan ahead and strategize in order to efficiently achieve the target configuration. Researchers commonly use the number of moves and the time taken to complete the task as indicators of learning, which are often interpreted as measures of implicit learning. There are also alternate versions such as the Tower of Hanoi (Simon, 1975) and Tower of Toronto (Saint-Cyr et al., 1988) which make minimal adjustments to the original task. Despite its use in assessing implicit learning, the task has been criticized for relying heavily on explicit processes such as planning and strategizing, as well as placing high demands on working memory, making its classification as a measure of implicit learning debatable.

2.2.8.7. ALTM synonym task

The ALTM Synonym Task is part of the Hi-LAB aptitude battery (Linck et al., 2013) and is designed to measure semantic or associative priming. It includes two components: the priming task and the comparison task. In the priming task, participants listen to a list of five words and are then asked to choose, from a pair of words, which one is synonymous with more words from the original list. One word in the pair is synonymous with two of the original five, while the other is synonymous with three. In the comparison task, participants are presented with word pairs and must indicate whether their meanings are similar or different. The final score, calculated using an elaborate method, reflects the level of priming or implicit learning, with higher scores indicating more priming. In a recent study by (Granena, 2019), factor analysis showed that the priming scores from the ALTM Synonym Task loaded together with LLAMA D, a subtest of the LLAMA battery that measures auditory recognition ability.

2.2.8.8. LLAMA D

LLAMA D (Meara, 2005; Meara & Rogers, 2019) is another measure often hypothesized to assess implicit aptitude (Granena, 2016, 2019; Iizuka & DeKeyser, 2023). It is a sub-test of the LLAMA battery and it measures auditory pattern recognition ability by asking participants to recognize and react to familiar sounds from an unknown language to which they were previously exposed. It is an auditory and non-visual measure that assesses the product of learning, i.e., accuracy (Christiansen, 2019). However, the status of LLAMA D as a measure of implicit aptitude remains questionable due to its inconsistent relationship with scores from the SRT (Iizuka & DeKeyser, 2023; Yi, 2018) and its lack of predictive validity for L2 proficiency (Suzuki, 2021).

2.3. Working memory

Working memory (WM) is a term first mentioned in Miller et al. (1960) and refers to a system responsible for the temporary storage and manipulation of information during the performance of higher-order cognitive tasks, such as comprehension, learning, and reasoning (Baddeley & Logie, 1999). Since L2 learning engages a variety of cognitive processes, it has been suggested that individual differences in WM capacity may be linked to variation in performance on cognitive and language tasks (Miyake & Shah, 1999; Oberauer et al., 2003). As a result, researchers in SLA have explored the extent to which such individual differences account for variability in L2 outcomes. To date, research has examined the role of WM in L2 reading (Alptekin & Erçetin, 2010; Leeser, 2007; Tyler, 2001; Walter, 2004), writing (Abu-Rabia, 2001; Adams & Guillot, 2008), sentence processing (Felser & Roberts, 2007; Juffs, 2004), speaking (O'Brien et al., 2006), grammar (Williams & Lovatt, 2003), and vocabulary development (Cheung, 1996; Papagno & Vallar, 1995; Speciale et al., 2004). Moreover, some scholars have proposed that WM may function as a component of language aptitude (Hummel, 2009; Kormos & Sáfár, 2008; Robinson, 2005, 2007, 2012; Skehan, 2016; Wen, 2019), highlighting its predictive relationship with overall L2 proficiency (van den Noort et al., 2006).

The subsequent sections provide an overview of the structure of WM and detail the most widely used measures to assess WM in L2 research. This is followed by empirical studies investigating its predictive power in relation to various aspects of L2 development. Finally, the review introduces the notion of L2 proficiency as a potential mediator of WM effects on language learning outcomes.

2.3.1. Structure of working memory

2.3.1.1. Componential view

Baddeley's structural model of working memory (Baddeley, 1986; Baddeley & Hitch, 1974) has been the dominant framework in both theoretical and applied research in SLA. In this classic fractionated model, working memory is composed of four components, two of which are modality-specific storage systems, while the fourth serves as a supervisory system that regulates the functioning of the other three. The phonological loop is responsible for processing spoken and written material. It consists of two subcomponents: the phonological store, which temporarily retains information in a speech-based form for approximately 1-2 seconds, and the articulatory control process, which can circulate information in a loop and converts written material into articulatory code and transfers it to the phonological store for further processing. The visuospatial sketchpad manages visual and spatial information, such as details about the appearance of objects and spatial orientation. It plays a crucial role in processing information about the surrounding environment and the relationship between objects within it. The episodic buffer, the third storage system, serves as a temporary holding space for integrating different types of information (e.g., verbal, visual, and spatial) and maintains a sense of time, allowing events to be experienced in a continuing sequence. These three components are often referred to as "slave systems," reflecting their passive role in storing limited amounts of information for brief periods of time, which is subject to constraints such as the chunk capacity limit and memory decay (Cowan, 2005). They are directly linked to long-term memory and are thought to serve as bottlenecks through which information must pass before being permanently stored (Juffs & Harrington, 2011).

The three storage systems are regulated by the central executive, which serves as the master component. Its functions include: (1) integrating information from multiple sources into a coherent episode (updating), (2) managing the shifts between task execution and the retrieval processes required for task completion (switching), and (3) controlling selective attention to focus on relevant information while inhibiting distractions (inhibition). Within the context of SLA, the control of attention has been considered the most critical function of the central executive (Juffs & Harrington, 2011).

2.3.1.2. Working memory and aptitude

In addition to Baddeley's componential view, other cognitive researchers advocate a more functional approach to working memory. For instance, the embedded-process model (Cowan, 1999, 2005) conceptualizes working memory as primarily a mechanism for attention control, while the executive control model (Engle & Kane, 2004) emphasizes the role of executive control.

In the Phonological/Executive (P/E) model (Wen, 2016; Wen, 2019), phonological working memory (PWM) and executive working memory (EWM) are considered key components of aptitude. PWM has been shown to play a critical role in both first language (Cogan et al., 2017; Pierce et al., 2017) and second language acquisition (N. Ellis, 1996; Foster et al., 2013; Wen, 2016), particularly among beginner- and intermediate-level learners (Serafini & Sanz, 2016). The core premise of the P/E model is that PWM functions as a modality-specific component primarily involved in language acquisition (Wen, 2016). At the same time, EWM is regarded as serving domain-general functions and processes such as switching, updating, and inhibition (Miyake & Friedman, 2012). EWM is thus crucial for cognitively demanding language processes in SLA, including sentence processing, discourse comprehension, production, and interactions, functioning as a domain-general component primarily involved in language processing (Wen, 2016).

Empirical investigations into the relationship between WM and language aptitude have yielded mixed results. Some studies have found no or only weak relationships between the two constructs (Roehr & Gánem-Gutiérrez, 2009b; Yoshimura, 2001). Li's (2016) meta-analysis, based on 66 studies including more than 13,000 L2 learners, suggested a significant correlation between EWM and explicit aptitude, but a weak or nonexistent relationship between PWM and explicit aptitude. Moreover, some findings suggest that PWM is particularly important for the acquisition of implicit knowledge, while EWM appears to be more closely related to the acquisition of explicit knowledge (Révész, 2012). A key limitation of previous research is the inconsistent operationalization of working memory, with some studies focusing solely on phonological or executive working memory rather than both. Similarly, most studies have not distinguished between implicit and explicit aptitude, and those that have often failed to account for the multicomponential nature of both. To fully understand the interrelations between these components, it is essential that studies measure all components within a single study.

2.3.2. Measures of phonological and executive working memory

Working memory capacity is determined by both storage and processing components. Simple short-term memory capacity is typically assessed by the number or span of unrelated digits or words that can be recalled. Processing capacity, however, is measured using tasks that demand both storage and processing, often referred to as complex working memory tasks. For both types of task, valid and reliable results depend on participants operating at optimal capacity and with minimal opportunity for strategic processing that might artificially enhance performance in ways unrelated to memory capacity (Juffs & Harrington, 2011). To ensure this, stimuli should be presented in such a manner that immediate responses are required.

2.3.2.1. Short-Term Memory Tasks

Short-term memory tasks typically involve word or digit span tasks, where participants recall sets of unrelated words or numbers presented either visually or aurally (Shah & Miyake, 1996). These stimuli are usually presented in ascending order of difficulty, with the number of digits or words increasing until the participant reaches their limit. When words are used, prior familiarity with the words may influence performance, thereby confounding PWM capacity and language knowledge (Gathercole, 1995). Therefore, digit span tasks are generally preferred in L2 research (Harrington & Sawyer, 1992). An alternative measure is a non-word repetition task, where participants are asked to repeat non-words that either follow the phonotactic rules of an existing language or include unfamiliar sounds. Performance on these tasks has been linked to word learning in both neurotypical children (Baddeley et al., 1998) and children with language impairments (Gathercole, 2006). It has also been associated with vocabulary acquisition (Service, 1992; Service & Kohonen, 1995; Speciale et al., 2004), oral production (O'Brien et al., 2006), and L2 proficiency (Hummel, 2009).

2.3.2.2. Complex Working Memory Tasks

The most widely used task to assess both storage and processing is the Reading Span task (Daneman & Carpenter, 1980). This task measures an individual’s ability to simultaneously read and comprehend a set of sentences and subsequently recall a target word from each sentence, usually the final word. Sentences are presented in sets, typically ranging from two to six sentences. To ensure that executive control is engaged, participants are also asked to perform a secondary comprehension task alongside recalling the target words. A spoken format, known as the Listening Span task, has also been developed, requiring participants to listen to sentences instead of reading them (Mackey et al., 2010; Mackey et al., 2002). However, comprehension tasks rely heavily on language knowledge, which can be mitigated by using alternative stimuli, such as arithmetic equations, instead of sentences (Turner & Engle, 1989). One such adaptation is the Operation Span task, where participants must recall target words while solving simple math equations following each sentence. To further minimize reliance on language proficiency, letters can be used instead of sentences (Conway et al., 2005).

Another alternative is the Backward Digit Span task (Kormos & Sáfár, 2008), in which participants hear sequences of spoken digits and are required to repeat them in reverse order.

This method significantly reduces the role of language knowledge, as digits are used in place of letters or sentences. In addition to reducing the reliance on linguistic processing, using L2 stimuli in these tasks is not ideal, as it may impose additional cognitive demands on working memory, potentially skewing results based on participants' proficiency levels. Any deficits in L2 knowledge could negatively affect performance (Sagarra, 2007, 2017). For these reasons, even tasks like the Backward Digit Span are best conducted in participants' L1.

2.3.3. Predictive power of working memory

A meta-analysis by Linck et al. (2014), which included 77 studies and over 3,700 participants, confirmed the crucial role of working memory in L2 processing and proficiency outcomes. Covariate analysis revealed that executive WM and verbal WM measures yielded larger effect sizes compared to phonological WM and nonverbal WM measures. This finding suggests that different components of WM (Baddeley, 2000) contribute distinctively to L2 processing, with executive functions such as updating, switching, and inhibition playing a key role above and beyond the simple maintenance of active representations in phonological short-term memory. Despite the prominence of Baddeley's (Baddeley, 1986; Baddeley & Hitch, 1974) componential model of WM, little is understood about how the three processes (updating, switching, and inhibition) relate to each other, aside from their separability (Miyake et al., 2000).

The meta-analysis further found larger effect sizes for L2 WM measures compared to L1 WM measures, though the researchers attributed this to overlap in the content of the predictor and outcome measures. An earlier study by Osaka and Osaka (1992) similarly found a strong relationship between L1 and L2 WM measures. However, a key limitation of the existing research on WM and L2 outcomes included in the meta-analysis is that it is largely correlational, making causal inferences difficult. There is some evidence indicating that executive WM can be systematically trained, resulting in improved performance not only on similar tasks (i.e. near transfer; Harrison et al., 2013; Sprenger et al., 2013) but also on language processing tasks that require executive control (i.e. far transfer; Novick et al., 2013). Thus, the findings suggest that WM training may enhance L2 processing abilities, supporting a potential causal link between WM and L2 outcomes (Linck et al., 2014).

Research focusing on particular L2 skills shows that L1 and L2 speakers use WM differently in reading, with L2 speakers relying more on topic knowledge for comprehension (Leeser, 2007; Tyler, 2001). While WM helps with inferential reading (Alptekin & Erçetin, 2010), L2 learners often depend more on top-down processing (Clahsen & Felser, 2006; Juffs & Harrington, 2011). In L2 writing, WM plays a crucial role, as higher WM capacity enhances performance across tasks like vocabulary, spelling, and composition (Abu-Rabia, 2001; Adams & Guillot, 2008).

Other studies have highlighted the role of PWM in L2 vocabulary learning, with Speciale et al. (2004) showing that both PWM and phonological sequence learning ability contribute to vocabulary acquisition in experimental and classroom settings. Their findings, replicated across two experiments, suggest that PWM is particularly important for productive vocabulary learning. Other studies also support the predictive role of nonword repetition tasks for vocabulary learning (Masoura & Gathercole, 2005; Papagno & Vallar, 1995; Service, 1992; Service & Kohonen, 1995), though some (Akamatsu, 2008; French & O'Brien, 2008) found no such relationship. Interestingly, Cheung (1996) noted that the effect of PWM is more prominent among lower proficiency learners, suggesting that L2 proficiency mediates PWM’s influence.

2.3.4. WM and L2 grammar

Of particular interest to the current study is research investigating the relationship between WM and L2 grammar learning. One of the few longitudinal studies examining this issue (Biedroń et al., 2022) collected data over a two-year period and explored the predictive roles of both PWM and EWM on grammar knowledge of Polish learners of L2 English. Grammar was assessed via multiple-choice tests, verb form filling, key word paraphrasing, and cloze tasks. PWM was measured using a combination of digit span and non-word span tasks, while EWM was assessed through reading and listening span tasks. All tasks used stimuli in the participants' L1. The predictive analysis revealed that only EWM, as measured by the listening span task, predicted grammar change over time. Correlation analysis indicated that both PWM and EWM were weakly to moderately related to grammar knowledge.

Another longitudinal study (Sagarra, 2017) investigated L2 Spanish learners across two experiments. Experiment 1, which spanned two semesters, found that EWM, measured using a reading span test (Daneman & Carpenter, 1980), did not significantly predict grammar scores, nor did it interact with the time variable. In Experiment 2, which spanned one semester, the sample size was increased, and the reading span test was replaced with a more demanding version (Waters & Caplan, 1996). While the linguistic tests remained the same, the results showed both a main effect of EWM and a significant interaction between EWM and time.

Among cross-sectional studies, Pawlak and Biedroń (2021) focused on L1 Polish learners of English. PWM was assessed using a non-word span test, while EWM was measured through a listening span test. Grammar knowledge was operationalized using both explicit and implicit tests focused on the passive voice in English. Receptive implicit knowledge was measured through a timed grammaticality judgment test, and productive implicit knowledge was assessed via a focused communication task. Receptive explicit knowledge was measured using an untimed grammaticality judgment test, while productive explicit knowledge was probed through a verb form completion task. The results showed that EWM significantly predicted explicit productive grammar knowledge, accounting for 6% of the variance, while PWM was related to both implicit and explicit productive knowledge of the passive, although it explained only about 2.6% of the variance. These findings confirm the greater role of EWM compared to PWM for L2 grammar knowledge.

In two studies by Kempe and colleagues (2008; 2010), complete beginner learners of Russian demonstrated that greater EWM, measured via a reading span test (Daneman & Carpenter, 1980), facilitated better connections between words and inflectional morphology, such as gender and case. Similarly, Sagarra (2007) found that beginner learners of Spanish, assessed using a self-paced reading test said to gauge implicit knowledge, were insensitive to gender violations, while learners with higher EWM, measured by the reading span test (Waters & Caplan, 1996), showed sensitivity to such violations.

Grey et al. (2015) examined morphosyntactic development in advanced learners of Spanish, operationalized as sensitivity to violations in gender and number agreement, as well as word order. Grammar knowledge was measured through a grammaticality judgment test, while PWM was assessed using both L1 and L2 non-word repetition tasks, and EWM via a sentence span test in L1. Although participants improved in their knowledge of number agreement and word order, neither PWM nor EWM was related to these changes. The authors argued that the richness of input in a study-abroad setting, combined with the participants' advanced proficiency, neutralized the effects of WM on L2 grammar learning.

Taken together, these studies seem to suggest that PWM and EWM may play distinct roles in grammar learning, with their influence potentially depending on the type of grammar knowledge being measured (implicit vs. explicit). The facilitative effects of WM on L2 grammar learning seem to be more pronounced among beginner learners (Kempe & Brooks, 2008; Kempe et al., 2010; Sagarra, 2007), whereas these effects may diminish or be neutralized among more proficient learners (Grey et al., 2015).

2.3.5. Effects of L2 proficiency

Studies directly investigating the role of L2 proficiency in mediating the effects of WM on L2 are limited. Several researchers have explored the relationship between WM and overall proficiency to determine the predictive role of WM in L2 learning. Kormos and Sáfár (2008) conducted a study on L1 Hungarian ESL learners, examining the predictive power of phonological working memory (PWM), measured via a non-word repetition task, and executive working memory (EWM), measured via a backward digit span task. A key finding was that scores on the two measures did not correlate, indicating a clear distinction between PWM and EWM. Surprisingly, PWM was not associated with any L2 proficiency measures (reading, writing, listening, or use of English) at the beginner level and was only weakly related to writing performance and the use of English at the pre-intermediate level. This result contrasts with previous research, which emphasized the importance of PWM in early L2 learning stages, particularly in lexical acquisition (Baddeley et al., 1998; Cheung, 1996). Meanwhile, EWM correlated with most test components, except for writing, and was linked to overall proficiency. The authors suggested that in classroom-based L2 instruction, PWM plays a limited role, while EWM has a more significant impact.

More interestingly, some researchers have acknowledged L2 proficiency as a mediating factor in the relationship between WM and L2 learning. Hummel (2009) conducted a study that examined the role of PWM in lower- and higher-proficiency groups. PWM was measured via a word span task, and L2 proficiency was assessed through vocabulary, grammar, and reading comprehension. Overall, WM predicted L2 proficiency, explaining 29% of the variance in proficiency when combined with language aptitude. However, when learners were divided into lower- and higher-proficiency groups, PWM accounted for 20% of the variance in the lower-proficiency group but only 5% in the higher-proficiency group, where it was not a significant predictor. This suggests that PWM is more relevant at lower proficiency levels.

The most compelling evidence for the mediating role of proficiency in the effects of WM on L2 learning comes from the longitudinal and cross-sectional study by Serafini and Sanz (2016), who included learners across a range of proficiency levels (beginner, intermediate, advanced). In their study, 87 students completed a digit span task to measure PWM and an operation span task to assess EWM. The outcome measures included ten morphosyntactic structures in L2 Spanish, tested via an elicited imitation task and an untimed grammaticality judgment task. The findings revealed that both PWM and EWM were most significant at the beginner and intermediate levels, while at the advanced level, WM played a minimal role in L2 development. The authors also suggested that, while it's undeniable that L2 proficiency mediates the effects of WM, the difficulty of the targeted morphosyntactic structures might serve as a moderating factor too: less experienced learners may rely more on WM when encountering challenging structures, whereas more advanced learners find these features less demanding, reducing their reliance on WM in tasks such as elicited imitation and grammaticality judgment tests.

2.4. Summary

Language learning aptitude has been identified as the strongest and most consistent predictor of L2 success, despite the heterogeneity in aptitude measures over the years (Li, 2015, 2016). The validity of this assertion, however, is contingent not only on the robustness of aptitude measures but also on the appropriateness of outcome measures. Methodological studies in this area have yielded less consistent results, particularly with certain measures such as timed grammaticality judgment tasks and elicited imitation, which have proven difficult to categorize as measures of either implicit or explicit knowledge (Godfroid & Kim, 2021; Suzuki & DeKeyser, 2015). Preliminary behavioural (Suzuki & DeKeyser, 2015, 2017) and neurocognitive evidence (Suzuki et al., 2023) suggests that these tasks measure explicit knowledge and automatized explicit knowledge, respectively, although limited evidence warrants further inquiry.

Moreover, several studies have highlighted the multi-componential nature of language aptitude, emphasizing distinct aptitudes for both implicit and explicit learning (Granena, 2013a; Li & DeKeyser, 2021; Linck et al., 2013; Roehr-Brackin et al., 2023). While the multifaceted nature of explicit aptitude has been well-established since the early days of aptitude research (Carroll, 1981), emerging evidence suggests that implicit aptitude is equally multifaceted (Granena, 2013a, 2020; Iizuka & DeKeyser, 2023; Li & DeKeyser, 2021; Roehr-Brackin et al., 2023). A major limitation, however, is the small number of reliable measures for implicit aptitude, which typically show low reliability, thus restricting the scope of empirical investigation. Additionally, existing studies suggest a lack of convergence among these measures (Godfroid & Kim, 2021; Iizuka & DeKeyser, 2023). The role of working memory as a predictor of grammar learning and as a part of language aptitude also remains unresolved. Although evidence clearly differentiates the roles of phonological working memory and executive working memory in L2 grammar learning (Biedroń et al., 2022; Pawlak & Biedroń, 2021), the heterogeneity of measures and discrepancies in findings make it difficult to determine when each component is most relevant or ceases to have an impact. Furthermore, the effects of WM components may vary depending on the type of grammar knowledge being measured, with only one study thus far showing this (Pawlak & Biedroń, 2021).

L2 proficiency has been identified as a potential mediating factor in the relationship between aptitude and L2 outcome. Studies that included L2 proficiency as a variable have indicated that explicit aptitude may play a more important role at lower proficiency levels but diminish as learners reach higher levels of proficiency (Artieda & Muñoz, 2016; Li, 2015). This mediating effect may also depend on the specific component of explicit aptitude in focus (Artieda & Muñoz, 2016). These findings align with theoretical models of aptitude, such as the staged (Skehan, 2002, 2012, 2016) and aptitude-treatment interaction models (Robinson, 2005, 2007, 2012). Additionally, there is theoretical support for the idea that implicit aptitude becomes more important at advanced proficiency levels (Li & DeKeyser, 2021), although direct empirical evidence is limited to a single study (Linck et al., 2013).

Working memory has similarly been shown to predict L2 success (Linck et al., 2014), albeit to a lesser extent than language aptitude. As with aptitude, the mediating role of proficiency extends to WM, with research indicating that WM is more influential at lower and intermediate proficiency levels (Grey et al., 2015; Sagarra, 2007; Serafini & Sanz, 2016). Furthermore, different components of working memory, such as phonological working memory and executive working memory, appear to play distinct roles depending on proficiency level (Hummel, 2009; Kormos & Sáfár, 2008).

Another key factor in the aptitude-L2 relationship is the difficulty of the linguistic structures being learned. Empirical evidence supports the notion that explicit aptitude plays a more important role when learners are confronted with more challenging structures (Robinson, 2002; Yalçın & Spada, 2016). In the case of WM, cumulative research suggests that EWM plays a more important role than PWM. More difficult linguistic structures are likely to place higher cognitive demands on learners, thereby activating EWM and underlying processes such as updating, switching, and inhibition. Such increased cognitive load could potentially amplify the significance of individual differences in EWM, while in the case of easier structures, this cognitive demand would not be triggered, rendering EWM (or WM) less- or non-significant.

Finally, the nature of the relationship between implicit and explicit knowledge has garnered increasing attention in recent years. Evidence suggests that an interface exists between the two types of knowledge. A cross-sectional study (Suzuki & DeKeyser, 2017) examining the relationship between language aptitude and L2 knowledge provided indirect evidence of an explicit-implicit interface. A more recent longitudinal study (Kim & Godfroid, 2023), however, demonstrated that the relationship can be bidirectional, with explicit knowledge influencing implicit knowledge and vice versa. Alternative theoretical accounts, however, propose the co-existence of multiple interfaces that vary depending on the complexity of the structure in question. While this approach could potentially explain the conflicting evidence observed thus far, further empirical support is necessary.

2.5. Research aims and questions

This review highlights several unresolved issues and overarching research questions that guide the studies reported here. First, it remains unclear whether implicit aptitude, like explicit aptitude, is multi-componential, and if so, which cognitive abilities constitute implicit aptitude. Additionally, in light of the lack of empirical support for considering WM as a part of language aptitude, it remains an open question whether WM should be operationalized as a separate construct or integrated into the framework of language aptitude. The present project, therefore, seeks to address these gaps by incorporating a range of measures of language aptitude and WM to investigate the structural properties of these cognitive abilities.

Second, while the predictive and explanatory roles of explicit aptitude and WM are relatively well-established, the role of implicit aptitude remains less understood. Furthermore, although some studies suggest that L2 proficiency may mediate the effects of explicit aptitude and WM, the available evidence is limited. A critical need exists to explore the role of implicit aptitude in conjunction with the potential mediating effect of L2 proficiency too. Moreover, rather than focusing on finding evidence that proficiency mediates the effects of aptitude and working memory, it would be valuable to pinpoint the exact point on the proficiency continuum where the shift from significant to non-significant, or vice versa, occurs for both aptitude and working memory. Similarly, it is unclear whether the moderating influence of structure difficulty, well-documented in the case of explicit aptitude, also applies to implicit aptitude and

WM. This project tackles these issues by using the constructs of explicit and implicit aptitude, alongside WM, to establish their predictive relationships with L2 knowledge. Proficiency is included as a mediating factor, while the targeted L2 structures represent grammatical features of varying difficulty, allowing for a comprehensive investigation of any interactions with aptitude and L2 knowledge.

Finally, if the mediating and moderating roles of L2 proficiency and structure difficulty are confirmed, it becomes essential to revisit the question of the interface between implicit and explicit knowledge, taking these factors into account. Thus, the project re-examines the issue of the interface between implicit and explicit knowledge by focusing on the relationship between language aptitude and L2 knowledge, employing a range of cognitive abilities to construct implicit and explicit aptitude. A variety of L2 knowledge measures are used, spanning from more implicit to more explicit. Specifically, implicit knowledge is assessed through a self-paced reading task, automatized explicit knowledge through elicited imitation test, and explicit knowledge via a gap-fill test. By considering the mediating role of L2 proficiency, the study aims to provide a more nuanced understanding of how these factors interact. The following questions guided the thesis:

2.5.1. The construct of language aptitude

RQ1: Is there evidence of convergence between auditory pattern recognition ability as measured by LLAMA D and implicit sequence learning ability as measured by a probabilistic SRT task?

RQ2: What is the relationship between measures of aptitude for explicit and implicit learning and measures of WM?

2.5.2. The predictive power of language aptitude and working memory at different levels of L2 proficiency

RQ3: To what extent do aptitude for explicit and implicit learning and WM predict L2 proficiency?

RQ4: To what extent are the effects of explicit and implicit aptitude and working memory on learners' knowledge of selected L2 morphosyntactic structures mediated by L2 proficiency?

RQ5: At which point on the L2 proficiency continuum do the facilitative effects of explicit and implicit aptitude and working memory become or cease to be significant?

2.5.3. Measures of explicit and implicit L2 knowledge

RQ6: How are scores on the self-paced reading, elicited imitation, and gap-fill tests related?

RQ7: How do explicit and implicit learning abilities contribute to learners' implicit knowledge?

RQ8: How do explicit and implicit learning abilities contribute to learners' explicit and automatized explicit knowledge?

2.5.4. The learner's perspective: structure difficulty, source attributions, and confidence

RQ9: Are the effects of explicit and implicit aptitude and working memory on learners' L2 knowledge of selected L2 morphosyntactic structures moderated by structure difficulty?

RQ10: What is the relationship between source attributions, confidence ratings, and accuracy on a gap-fill test?

3. Methodology

This dissertation aims to uncover the complex interplay of variables underlying implicit and explicit language aptitude and working memory, with particular emphasis on the potential multicomponential nature of implicit aptitude. The impact of these individual differences on language outcomes is also examined, alongside the potential mediating role of L2 proficiency and the moderating role of structure difficulty. Additionally, the interface between implicit and explicit knowledge is re-examined by scrutinizing the relationship between aptitude for implicit and explicit learning and the implicit, explicit, and automatized explicit knowledge.

The research questions were investigated using a correlational study that employed the LLAMA battery and the serial reaction time task to measure explicit and implicit aptitude, and digit span and operation span tasks to assess working memory. Language outcomes were assessed through a self-paced reading task, a gap-fill test, and an elicited imitation test, which measured implicit, explicit, and automatized explicit knowledge, respectively. Proficiency was evaluated using the Oxford Placement Test and an oral production task. The following sections will provide details about the participants, a comprehensive description of all the instruments and stimuli, as well as the results of the pilot study. It will also outline the changes made to the design of the main study based on these pilot results.

3.1. Participants

The study sample comprised 86 English-as-a-Foreign-Language students enrolled in the first through fourth grades of three *public gymnasium* high schools (grammar schools) in Croatia, aged between 15 and 18 years (M = 16.14, SD = 1.29). Participants had been learning English for 6 to 13 years (M = 10, SD = 1.72) as part of their mandatory school curriculum. The sample included 62 females, 22 males, and 2 participants who preferred not to disclose their gender.

All participants had been studying English as a compulsory subject, receiving three hours of instruction per week since the beginning of high school. Their English classes focused on both communicative language skills and form-focused instruction to prepare for the nationwide exam required at the end of their final year. According to the Ministry guidelines and the national curriculum, students at this stage are generally expected to reach at least A2 to B1 proficiency levels (Medved Krajnović, 2007), as defined by the Common European Framework of Reference (CEFR). However, based on communication with the students' language teacher (D. Linić Učur, personal communication, June 7, 2020), it was anticipated that these students would perform slightly above the national average, in part due to the high entry requirements of their secondary schools. Supporting this expectation, a recent global survey by Education First ranked Croatian students fifth worldwide in English language proficiency (Education First, 2024).

The rationale for selecting this learner profile was twofold: first, participants needed to be sufficiently proficient to complete the language assessments successfully and to allow for the detection of implicit aptitude effects, as suggested by previous research (Linck et al., 2013), and second, it was important to ensure that the sample also included learners of lower proficiency, in order to allow for discrimination between lower- and higher-proficiency learners.

The majority (78%) used *Pearson's Focus* as their English coursebook, while some (17%) used *Oxford's Insight*, and a minority (5%) used *Pearson's High Note*.

In addition to English, all participants were enrolled in language classes of their choice as part of their school curriculum. Of the participants, 48 (56%) were taking German, having studied it for an average of 5.2 years. Italian was the second most popular choice, with 34 participants (40%) studying it for an average of 5.76 years. French was the least popular, with only 4 participants (4%) choosing it and having studied it for an average of 3 years.

Regarding exposure to English-speaking environments, 70 participants (81.4%) had never visited an English-speaking country, while 2 participants (2.3%) had stayed in such a country for a year, and the remaining 14 participants (16.3%) reported short stays averaging 25 days.

3.2. Target structures

Objective determinants of difficulty refer to the properties of the target structure. Roehr and Gánem-Gutiérrez (2009a), DeKeyser (2005), and R. Ellis (2006) suggest defining the difficulty of a particular L2 grammar feature by evaluating its characteristics against a set of criteria. As detailed in Section 2.2.7, these criteria include frequency, perceptual salience, communicative redundancy, opacity, schematicity, conceptual complexity, the technicality of metalanguage, and truth value. Subjective difficulty, on the other hand, depends on the learner's profile and cross-linguistic differences, which account for whether a structure exists in an L1 or how differently it is used across languages. In addition to the descriptive perspective, Collins et al. (2009) proposed incorporating a pedagogical approach which takes into account how these features align with the instructional context. Using these frameworks, three grammatical features of English were selected and assessed for their relative difficulty.

3.2.1. Articles

Articles are used in English noun phrases to indicate the identifiability of referents and to specify the grammatical definiteness of noun phrases. The English article system comprises the indefinite article "a(n)", the definite article "the", and the zero (or "null") article (Hawkins, 2015). This system is widely considered one of the most challenging structural elements for ESL learners to master (Liu & Gleason, 2002). To illustrate its difficulty, consider the following examples from Hawkins (2015):

A.1 *Fred was discussing <u>an interesting book</u> in his class.*

A.2 *I went to discuss <u>the book</u> with him afterwards.*

In these two sentences, a first-mention indefinite description (*an interesting book*) is followed by a second-mention definite description (*the book*). It is clear that *the book* in A.2 refers to the same object as the preceding indefinite reference, and that only one object is being referenced in each sentence. Another example demonstrates a different use of the definite article:

A.1 *Fred was discussing <u>an interesting book</u> in his class.*

A.3 *He is friendly with <u>the author</u>.*

In A.3, there is no preceding reference to *an author*, but *the author* is understood as referring to the author of the previously mentioned book. This usage indicates that the definite article *the* links *the author* to the book previously described. In contrast, an indefinite article can be used in the same context to convey a different meaning:

A.1 *Fred was discussing <u>an interesting book</u> in his class.*

A.2' *I went to discuss <u>a book</u> with him afterwards.*

In A.2', the indefinite article *a* indicates that the book referred to is not the same as the *interesting book* mentioned in A.1. Similarly:

A.1 *Fred was discussing <u>an interesting book</u> in his class.*

A.3' *He is friendly with <u>an author</u>.*

In A.3', *an author* would most naturally be interpreted as someone who is not the author of the previously mentioned book, otherwise *the author* would have been used. These examples illustrate high opacity (single form mapped onto multiple functions), low perceptual salience (cannot be easily perceived in spoken input), and partial communicative redundancy (not crucial for the communicative intent of a message), thereby increasing the learning difficulty of acquiring implicit knowledge of articles. Furthermore, the multiple contexts and uses of articles indicate low schematicity (many examples for specific use) which increases both implicit and explicit learning difficulty. So, when judging by the objective difficulty, articles are difficulty for both implicit and explicit learning.

In comparison, the Standard Croatian language expresses definiteness in different ways, but it often cannot or is not explicitly marked. When definiteness is expressed, it is typically done using adjectives. For masculine gender nouns, definiteness can be marked through definite versus indefinite adjectives:

A.4 *Crn oblak. — A black cloud.* [Nominative]

A.4' *Crni oblak. — The black cloud.* [Nominative]

In this example, the adjective *crn(i)* marks the definiteness of the noun. However, for feminine gender nouns, the same form of the adjective is used for both definite and indefinite references, making it impossible to express definiteness this way. Similarly, for neuter gender nouns, different forms that allow expression of definiteness exist only in some cases (e.g., genitive, dative, locative) but not in others (e.g., nominative, accusative). Further complicating matters, Croatian speakers often omit definiteness markers when using declensions, and the definite form of the adjective tends to be preferred in both contexts, even when distinct forms exist (Pranjković, 2000).

Moreover, Croatian speakers frequently rely on secondary methods to express definiteness, such as using demonstratives (e.g., *this*, *that*) or numbers (e.g., *one*, *some*). Although definiteness can be conveyed in Croatian using demonstrative pronouns (e.g., *ovaj* [this], *onaj* [that], *takav* [such]), or indefinite pronouns (e.g., *neki*, *poneki*, *nekakav* [some]), or even the number *jedan/jedna/jedno* [one], the grammatical category of definiteness is often overlooked in Croatian grammar books (Zergollern-Miletić, 2008). As a result, most Croatian speakers are unaware of its presence in their own language and fail to recognize markers of definiteness when they are used, complicating the acquisition of the English article system by L1 Croatian learners.

Thus, there is very little correspondence in how definiteness is operationalized in the Croatian language compared to the English language. Specifically, the limited markedness of definiteness and the high optionality of those existing markers of definiteness in Croatian means that learning the English article system also requires raising learners' awareness of the grammatical category of definiteness. This significantly increases the subjective difficulty of learning articles. So according to criteria for both objective and subjective difficulty, the articles are considered to be difficult to learn both implicitly and explicitly for Croatian learners of English. For these reasons, we consider the article system to be a difficult structure within the scope of this study.

3.2.2. Simple past tense

The simple past in English is a form used to describe events that occurred in the past. It can be expressed through regular inflection, where the morpheme -ed is added to the base form of the verb (e.g., *walk – walked*). This pattern applies predictably to thousands of verbs and is even used for the formation of neologisms, such as spam – spammed or mosh – moshed (Pinker & Ullman, 2002). In contrast, irregular inflection (e.g., make – made or feel – felt) is applied to only around 180 verbs and does so unpredictably and is rarely generalized (Pinker & Ullman, 2002). Consequently, L2 learners of English must memorize irregular verbs which is facilitated by the fact that many of the irregular verbs are highly frequent.

In Croatian, the perfect tense is almost exclusively used to describe past events and does not entail any irregularities. It is a compound tense, formed using the present form of the verb to be ("biti") and an active verbal adjective that also encodes the gender of the person performing the action (Anđel et al., 2000). The English simple past tense and present perfect tense are both expressed using the same perfect tense in Croatian, as illustrated by the following examples:

B.1 Danas nisam ručao. — I did not have (any) lunch today.

B.2 Danas nisam ručao. — I have not had (any) lunch today.

A study on 135 secondary school students (Bagarić, 2001) found that differentiating between the past simple and present perfect is among the most challenging grammar points for Croatian students of English. This observation was further corroborated through personal communication with three EFL teachers in Croatia, each with more than 10 years of experience working with high school students (July 2020). The difficulty is particularly pronounced when considering that Croatian speakers do not need to differentiate between past actions based on definiteness or time expressions such as yesterday or last year, which in English determine the appropriate past tense to use in a given context.

Because of these cross-linguistic differences, Croatian learners of English receive extensive formal training on the use of the past simple tense. From a pedagogical perspective, EFL teachers consider the past simple a less challenging structure compared to articles. From the descriptive perspective, it has higher perceptual salience, especially in the case of irregular verbs, it has low communicative redundancy, and it is less opaque, or in other words, its form-meaning mappings are more transparent. Therefore, although the simple past tense may pose a challenge due to L1-L2 distance, it is still regarded as less complex than articles. For this reason, the simple past tense is considered to be on the easier side of the difficulty continuum, yet still challenging enough to prevent ceiling effects in testing and produce sufficient variation to discriminate between learners.

3.2.3. Passive voice

The passive voice is a grammatical construction in which the grammatical subject is the theme or patient of the main verb (O'Grady et al., 2001). In a passive sentence, what is typically expressed by the object of the verb is conveyed by the subject, while what is normally expressed by the subject is either omitted or expressed through an adjunct of the sentence. The use of the passive voice allows speakers to organize discourse by placing elements other than the agent in the subject position. This positioning may be employed to foreground the patient, recipient, or another thematic role (Saeed, 1997) or when the semantic patient is the topic of an ongoing discussion (Croft, 1991). In formal discourse, the passive voice is frequently used, yet it is often perceived as overly formal and verbose by students of English (Čupić & Klanjčić, 2015). In English, passive constructions are formed with the verb *be* followed by a past participle.

In Croatian, the passive voice is constructed similarly, using the verb to be ("biti") and a passive verbal adjective, and it can be formed with most transitive verbs:

C.1 Netko je očistio prozore. – Someone has cleaned the windows. [active]

C.1' Prozori su očišćeni. – The windows have been cleaned. [passive]

Although the passive has the same syntactic and functional properties in both languages, the Croatian language generally favours the active structure in both informal and formal contexts. This preference makes Croatian learners of English less inclined to use the passive and may partly explain their difficulty in adopting this feature when learning English (Čupić & Klanjčić, 2015).

In terms of objective difficulty, the passive voice is considered more complex than articles or the simple past tense because it introduces a higher cognitive load, particularly when used in combination with more complex tenses such as the present perfect or past perfect continuous. Additionally, from the perspective of implicit knowledge, the passive voice is more opaque compared to the simple past tense with relatively high communicative redundancy. Although the high correspondence between the English and Croatian passive might suggest an advantage, this benefit is somewhat diminished due to the minimal use of the passive voice in Croatian. Lastly, compared to the article system, schematicity for the passive voice is high which makes it easier to learn both implicitly and explicitly.

On the other hand, students enrolled in courses such as English for Academic Purposes (EAP) or English for Special Purposes (ESP) are more likely to receive extensive training on the use of the passive voice, as these programs are designed to prepare learners for formal or specialized language use. Similarly, mandatory English classes in Croatian secondary education have increasingly emphasized the use of the passive voice due to the introduction of national standardized tests of English 15 years ago, which prioritize written skills and academic-style essay tasks as crucial components of the final score. Consequently, while the passive voice is considered more difficult than the simple past tense, it is still regarded as less challenging compared to the English article system.

3.2.4. Typical errors and distractors

Both the self-paced reading test and the elicited imitation test required ungrammatical sentences alongside grammatical ones, while the gap-fill test required two distractors for each item. Below is further information describing the themes of errors, accompanied by example sentences illustrating each theme, while the full list of items can be found in Appendix 1. Sentences involving articles were constructed to reflect one of five error types:

- "A/an" with singular countable nouns (e.g. Roger works for firm in Midtown Manhattan.)

- Zero article when talking about things in general (e.g. Her coat is made of the pure wool.)
- "The" when there is only one of something (e.g. People used to think earth was flat.)
- "The" with plural names of people and places (e.g. It is said that Amazon river is one vast highway.)
- "The" to talk about a type of animal, instrument etc. (e.g. My sister has been playing violin for three years.)

Sentences involving the passive voice were designed to reflect errors in the supplied verb form belonging to a different tense:

- Present simple (e.g. How is this word use in a sentence?)
- Past simple (e.g. The movie ET was direct by Steven Spielberg.)
- Present perfect (e.g. The date of the meeting has been change and everyone was notified soon after.)
- Past continuous (e.g. Tom was being questioning at the police station when I Called him.)
- Infinitive (e.g. Kayaks can be rent at various shops around the island.)

Sentences involving past tense were designed to reflect one of four types of errors:

- Double marking (e.g. Did it rained on Sunday morning?)
- Present simple form (e.g. My father apply for this job four times before they called him back.)
- Present continuous form (e.g. The police stopping me on my way home last night.)
- Additional copula (e.g. My family was rented a villa in France every summer.)

3.3. Instruments

To measure explicit aptitude, LLAMA B assessed associative memory, LLAMA E measured phonetic coding ability, and LLAMA F evaluated language-analytic ability. Implicit aptitude was operationalized through auditory recognition ability, measured by LLAMA D, and implicit sequence learning ability gauged via the serial reaction time task. Implicit knowledge was measured using self-paced reading and word monitoring tasks, while automatized explicit knowledge was assessed through elicited imitation, and explicit knowledge via a gap-fill test. The following section will provide detailed information on these measures.

3.3.1. Self-paced reading task

The self-paced reading (SPR) task uses a moving window presentation, in which participants are instructed to read sentences as quickly as possible, one word at a time, while focusing on meaning in order to accurately answer subsequent comprehension questions. Participants press a key to move through the words in a sentence. This setup enables the recording of reaction times for each key press, and it requires participants to read from left to right, word by word, in a manner that closely resembles natural reading. Because each word disappears as the subsequent word appears, participants cannot review the entire sentence at any point.

In a study by Pearlmutter et al. (1999), it was shown that L1 English speakers are sensitive to grammatical violations, as indicated by delays in their reading of ungrammatical sentences, which were reflected in their reaction times recorded through key presses. The self-paced reading task has since been demonstrated to effectively capture implicit language processing in both L1 and L2 (Jiang et al., 2011; Roberts & Liszka, 2013).

3.3.1.1. *Instructions and presentation*

The stimuli were displayed against a white background. The instructions outlined the format of the task. Participants were informed that they would be reading the sentences for comprehension. Before each sentence, a cross symbol was displayed as a fixation point. The first word of each sentence appeared on the left side of the screen, and upon pressing a designated key, the next word appeared to the right, replacing the previous one. After each sentence, a comprehension question was presented to ensure that participants were focused on the meaning. The dual-task nature of this paradigm has been shown to minimize the reliance on explicit knowledge and strategy use (Kilborn & Moss, 1996).

3.3.1.2. *Stimuli and scoring*

The task included four regions of interest (ROIs): (1) the word preceding the grammatical violation (before the violation becomes apparent), (2) the critical word where the grammatical violation occurs, (3) the word immediately following the critical word, and (4) the word following the word at ROI 3. ROI 1 served as a baseline, ROI 2 was expected to show a delay in reaction time (RT), whereas ROIs 3 and ROI 4 were included to capture any spillover effects (Jiang, 2007). Table 5 shows example sentence for each of the three structures containing a grammatical error, highlighting all four regions of interest.

Table 5. Example sentences in self-paced reading task

Structure		ROI 1	ROI 2	ROI 3	ROI 4	
Past	When I was a child,	I	visit	my	grandma	every weekend.
Passive	The garbage	is	collect	every	day.	
Articles	People used to think	an	Earth	was	flat.	

Participants were unaware that some sentences were grammatical, while others contained errors. Delays in reaction time at ROIs 2, 3, and 4 in ungrammatical sentences were interpreted as indicators of the automatic activation of implicit L2 knowledge (Marslen-Wilson & Tyler, 1980). A grammatical sensitivity index (GSI) was calculated as the difference in reaction times between grammatical and ungrammatical sentences across ROIs 2, 3, and 4 combined (Suzuki, 2017).

The stimuli comprised 72 target sentences and 24 filler sentences. There were 24 sentences per target feature, with half of each set grammatically correct and the other half grammatically incorrect. All filler sentences were grammatically correct. Two counterbalanced test lists were created from the 72 sentences and their ungrammatical counterparts, ensuring that no sentence was repeated within the same list, and that the grammatical and ungrammatical versions of each sentence appeared in different lists. To reduce the possibility of raising awareness of the task's purpose or the specific target structures, target and filler sentences were interspersed and presented in random order. Sentences from the same structure never appeared more than twice consecutively.

Each sentence was followed by a yes/no comprehension question, evenly distributed between positive and negative responses. The comprehension questions were used to calculate overall comprehension accuracy, with 80% set as a cut-off point. Any participant scoring below this threshold would have been excluded from further analysis; however, no participants were excluded based on this criterion.

3.3.2. Elicited imitation with word monitoring component

The elicited imitation (EI) task is an online measure often suggested as a measure of implicit language knowledge (R. Ellis, 2005; Erlam, 2006; Godfroid & Kim, 2021). As discussed in Section 1.2.1.2. above, there is now growing behavioural and neurolinguistic evidence suggesting that the elicited imitation test primarily taps into automatized explicit knowledge. In an elicited imitation test, participants are asked to (a) listen to a stimulus sentence, (b) respond to a brief comprehension question, and (c) repeat the sentence they have heard. To make the task more sensitive to implicit knowledge, Suzuki and DeKeyser (2015) proposed incorporating a word monitoring component (Granena, 2012, 2013b) into the EI task, leading to the development of what is now referred to as an EIM task.

The word monitoring task (Marslen-Wilson & Tyler, 1975, 1980) is an online measure of language processing and can be used to gauge implicit L2 knowledge (Marslen-Wilson & Tyler, 1980; Kilborn & Moss, 1996; see Jiang, 2004, 2007, for a rationale with a similar procedure). In this task, participants read a monitoring word on the screen, listen to a sentence containing that word, and press a designated key as soon as they hear it. Participants are unaware that some sentences they hear contain grammatical errors, while others do not. These grammatical errors always appear immediately before the monitoring word. Thus, by measuring participants' reaction time to the monitoring word, any delay or hesitation caused by additional processing of the error is interpreted as evidence of implicit knowledge.

According to the EI design proposed by (Erlam, 2006), EI tasks typically include comprehension questions to shift learners' focus from form to meaning. However, the newly integrated word monitoring component already directs learners' attention away from form, as participants must listen attentively for the monitoring word and react as quickly as possible. Therefore, the version of the task in this study did not include comprehension questions. This also eliminated the need for semantically implausible filler sentences, whose primary purpose is to shift attention from form and elicit negative responses on comprehension questions. In sum, removing comprehension questions simplified the task for participants, while the inclusion of the word monitoring component maintained their focus on meaning (Suzuki & DeKeyser, 2015).

During the task, participants were first presented with a monitoring word at the centre of the screen and were instructed to press a designated keyboard button as soon as they heard the monitoring word in the sentence. The auditory sentence began 2 seconds after the monitoring word appeared and remained on the screen until a response was given. After the auditory sentence ended, a series of five random numbers appeared on the screen at one-second intervals, and participants were instructed to read the numbers aloud in Croatian as they appeared. This step was implemented to prevent rehearsal and rote repetition of the sentences (Mackey & Gass, 2022). Notably, this study utilized five numbers instead of the three used in previous studies employing the EIM task (Suzuki & DeKeyser, 2015). Additionally, the numbers were randomly selected for each trial, requiring participants to direct their attention to the numbers as they read them aloud. This increased cognitive load was designed to further compensate for the exclusion of comprehension questions. After this step, participants were shown an image of a microphone as a cue to begin sentence repetition. They had to complete the imitation of the sentence within nine seconds. This time limit was determined based on a pre-pilot study, which found that nine seconds was neither too short nor too long for participants to adequately repeat the sentence. The disappearance of the microphone icon marked the end of the response time, and participants had to press a designated keyboard button to proceed to the next sentence.

3.3.2.1. Instructions and presentation

The EIM instructions informed participants to: (a) press the button as soon as they heard the target word in the sentence, (b) read aloud the numbers as they appeared on the screen, and (c) repeat the sentence into the microphone, using different words if necessary, as long as the meaning remained unchanged. Participants were also instructed to repeat the sentences in grammatically correct English. This approach, which balances guidance with flexibility, aligns with methods used in previous studies (Erlam, 2006).

To ensure participants understood the task, a practice phase with six sentences was conducted. Three of these sentences contained grammatical errors, such as incorrect number agreement ("[…] three dog and one cat.") or incorrect tense usage ("I am go to […]"). After participants repeated each practice sentence, they were provided with the correct response for comparison. The practice phase also included dynamic, context-sensitive instructions displayed in the top-right corner of the screen. Instructions relevant to the current task phase appeared in black, while irrelevant instructions were shown in grey to help participants become familiar with the procedure and order of each step (see Figure 2).

Figure 2. Participant instructions during test phase of elicited imitation with word monitoring task

3.3.2.2. Stimuli and scoring

The stimuli consisted of 72 items representing the three target structures (three sets of 24 sentences), with each set containing an equal number of grammatically correct and grammatically incorrect sentences. These sentences were interspersed to ensure that no items testing the same structure appeared consecutively more than twice. Two counterbalanced lists were created: in List 1, the 36 grammatical sentences had corresponding ungrammatical versions in List 2, and vice versa. The sentences varied in length, ranging from 8 to 24 syllables, with an average length of 15.68 syllables. The study employed both simple and complex linguistic structures, and length is inherently tied to some of the targeted grammatical structures (e.g., the passive voice). Thus, the items represented a range of difficulty for participants, incorporating "stimuli of various lengths and complexities," as recommended by Bley-Vroman and Chaudron (1994) . In terms of duration, mean sentence length was 5.53 seconds, with the shortest sentence lasting 3.9 seconds. This exceeds the 1.5 to 2.0 seconds it typically takes for information to decay from phonological short-term memory without rehearsal or refreshing (Baddeley et al., 1975).

The elicited imitation scores were calculated following the conventions outlined in Erlam (2006) and Suzuki and DeKeyser (2015) based on the following categories: (a) obligatory occasion created – required form supplied, (b) obligatory occasion created – required form not supplied, and (c) no obligatory occasion created. A point was given only when the response fell into the first category. Responses falling into the second and third categories were considered incorrect and received no points. If an error was present in the response but did not involve the target structure, a point was still awarded, provided that the target structure was used correctly. Additionally, if participants self-corrected within the 9-second time limit, a point was still granted.

The position of the monitoring word varied across sentences, appearing after an average of 6.81 words, with a maximum of 13 and a minimum of 3 words preceding it. Speech rate in native American English has been shown to be around 150 words per minute (w.p.m.), depending heavily on context (Virtualspeech.com). Everyday conversation tends to have the highest speech rate, whereas academic lectures and presentations have the lowest. In this study, the speech rate was set at 123 w.p.m. This slower rate was deemed appropriate given that the participants were L2 English speakers, and the sentences were presented in isolation, which differs significantly from real-life language use. Scoring on the word monitoring task was calculated as the difference between reaction times between grammatical and ungrammatical sentences.

3.3.3. Gap-fill test

The gap-fill test included sentences with a gap and three provided options, only one of which was correct. The test had no time pressure and it aimed to engage participants' focus on form (R. Ellis, 2005), it was expected that participants would rely on their explicit knowledge to answer the questions. The test consisted of 75 sentences, equally divided among the three target structures:

- *Articles.* The three options provided for each test item were always (1) a(n), (2) the, and (3) zero (null) article. One option was correct, while the other two were incorrect.
- *Past Simple.* Distractors were created by considering several factors. Based on the findings from bin Abdullah (2013), the most common error type for L2 learners of English is addition due to double marking (e.g., “I didn’t realized […]” or “Did you saw […]”). When possible, this type of error was embedded in the first distractor. If that was not feasible, the present simple or past continuous form of the verb was used instead. The second distractor always utilized the present perfect tense, as Croatian L2 learners of English often struggle to distinguish between its use and that of the past simple (see Section 2.2.2.).
- *Passive Voice.* One distractor always involved an active form of the verb in the corresponding tense (e.g., Othello was written by […] -> Othello wrote by […]). The second distractor consisted of errors such as the wrong tense (e.g., have been solved -> solved) or the incorrect participle form of the correct tense (e.g., was sent -> was send).

These distractors were carefully constructed by taking into account the most frequent errors made by participants during their regular English classes, as reported by their English teacher. Each correct response was awarded one point, with a maximum possible score of 75 points.

3.3.3.1. Instructions and presentation

The certainty scale and source attributions were included as subjective measures of awareness (Maie & DeKeyser, 2020; Rebuschat et al., 2015). Upon selecting an answer, a certainty scale appeared on the screen, prompting participants to indicate their confidence level, ranging from 50% to 100%. This choice was informed by pilot study results (see Section 3.4), which showed that no participant selected a value below 50%, and it was theoretically justified since 50%

represents chance or guessing. Following the certainty scale, participants were asked to indicate the source of their response by selecting one of the following categories: rule, intuition, guess, or memory. The instructions explained each category in layman's terms to ensure comprehension by all participants. The descriptions used for each category were adapted from Tomak (2019) and are provided below:

- GUESS: You are guessing the answer.
- INTUITION: You think the answer is right, but you don't know why.
- MEMORY: You have a memory of having heard or encountered something similar.
- RULE: You know the rule and used it to answer.

3.3.4. LLAMA battery

The LLAMA battery (Meara, 2005; Meara & Rogers, 2019)is a language aptitude test loosely based on the MLAT created by Carroll and Sapon (1959). It aims to measure the key aptitude components of phonetic coding ability, language-analytic ability, and memory (see Section 2.2.3.6). The battery comprises four subtests, as follows:

LLAMA B is an associative memory task that assesses one's ability to quickly learn new vocabulary. In the learning phase, participants have 2 minutes to learn 20 vocabulary items from an invented language, each associated with a unique pictorial stimulus. By clicking on a pictorial stimulus, a corresponding word appears, and participants can click on the items as many times as they like. In the test phase, participants are presented with a word and must select the corresponding pictorial stimulus from 20 options. Both phases can be seen in Figure 3. There is no time limit during the test phase, and the maximum score is 20.

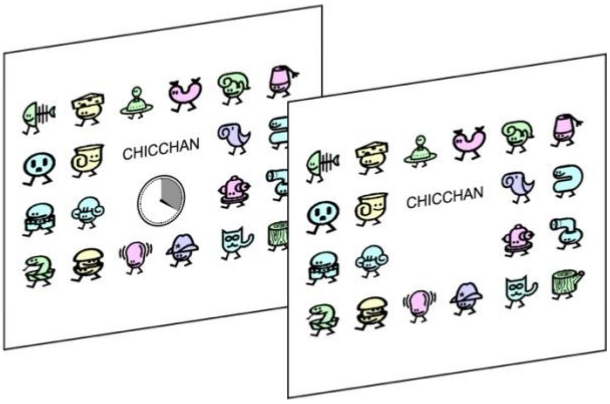

Figure 3. LLAMA B interface

LLAMA D is a sound recognition task that gauges auditory pattern recognition ability. In the exposure phase, participants listen to 10 words in an unknown language. The words are based on natural objects and flowers from a British Columbian indigenous language, synthesized with a French accent. In the test phase, participants hear a larger set of 40 words, including both familiar and unfamiliar ones. For each word, participants must indicate whether they have heard the word before or if it is new. Exposure and test phases can be seen in Figure 4. The maximum score is 40, with incorrect answers yielding negative points.

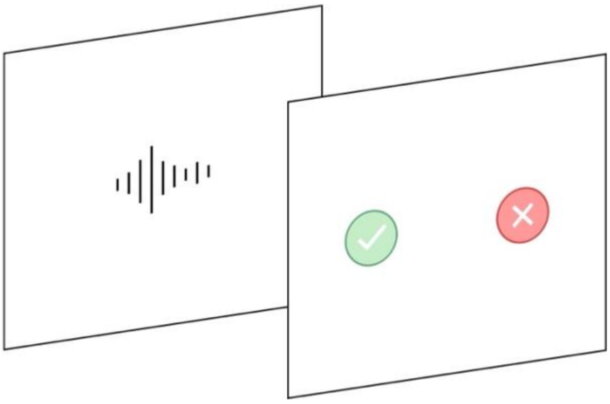

Figure 4. LLAMA D interface

LLAMA E is a sound-symbol correspondence task designed to assess phonetic coding ability. Participants are introduced to 22 symbols, each representing a unique syllable. By clicking on a symbol, the corresponding syllable is played. Participants can click on the symbols as many times as they like within the 2-minute learning phase. In the test phase, which comprises 20 items, participants hear combinations of previously learned syllables and must choose the correct symbol combination from 20 options (see Figure 5). The maximum score is 40, and participants can earn partial credit if they correctly select at least one of the two syllables.

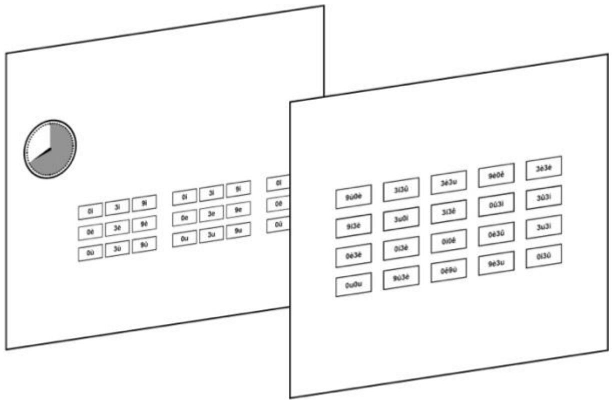

Figure 5. LLAMA E interface

LLAMA F is a grammatical inferencing task where participants have 4 minutes to deduce the rules of an unknown language. In the learning phase, participants are presented with a pictorial stimulus accompanied by a written description in the unknown language. There are 20 such items, and participants can click on different buttons to switch between stimuli as often as they wish. In the untimed test phase, participants are shown slightly different pictorial stimuli and must construct the correct written description from a range of words (see Figure 6). The maximum score is 132, and partial points are awarded – one point if the answer contains the correct word, and an additional point if the word is in the correct position.

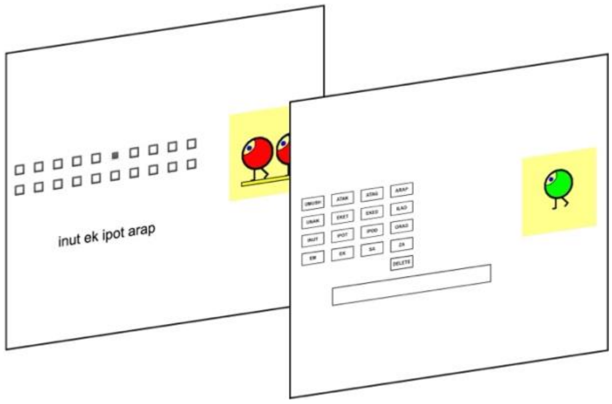

Figure 6. LLAMA F interface

3.3.5. Serial reaction time task

To gauge aptitude for domain-general implicit learning, a probabilistic serial reaction time (SRT) task was administered (Kaufman et al., 2010; Nissen & Bullemer, 1987). In the SRT task, participants see a stimulus that appears in one of four locations on the computer screen and are instructed to press the corresponding key as quickly and accurately as possible. Figure 7 displays the user interface.

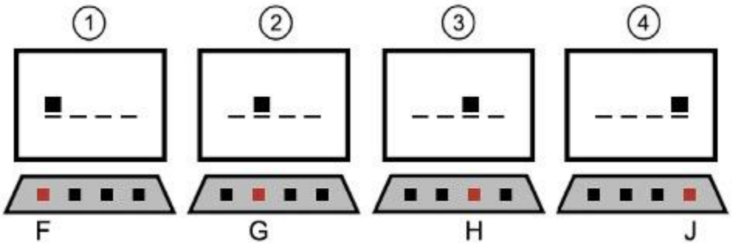

Figure 7. Serial reaction time task interface

The sequence of stimuli follows a probabilistic rule – 85% of the time, the sequence follows this rule (training condition), while the remaining 15% of the time, the stimuli follow a different sequence (control condition). Specifically, Sequence A (1-2-1-4-3-2-4-1-3-4-2-3) occurs with a probability of 0.85, and Sequence B (3-2-3-4-1-2-4-3-1-4-2-1) occurs with a probability of 0.15 in each block. The probabilistic nature of the task has two main advantages: it makes it less likely for participants to explicitly discover the existence of a sequence and its pattern, and it increases ecological validity, since implicit learning in real-world settings often depends on information that is noisy and probabilistic rather than deterministic (see Section 2.2.8.1).

A key attribute of the probabilistic SRT is that the two sequences are composed of entirely different second-order conditionals, meaning they cannot be predicted by first-order conditionals (Reed & Johnson, 1994). A first-order conditional sequence is determined by only the previous location, so the probability of the next cue's appearance at any given location is the same for both sequences – if the current stimulus is 1, the following stimuli can be either 2, 3, or 4 and this is true for both sequences. In contrast, a second-order conditional sequence is determined by the previous two locations, meaning the probability of the next cue can be unique to each sequence. For instance, in Sequence A, a cue of 1 followed by 4 is always succeeded by 3 (1-4-3), while in Sequence B, 1 followed by 4 is always succeeded by 2 (1-4-2). This distinction increases the task's complexity and promotes the implicit abstraction of probabilities while minimizing the potential for chunk learning (Suzuki & DeKeyser, 2015). Figure 8 displays possible transitions between sequences.

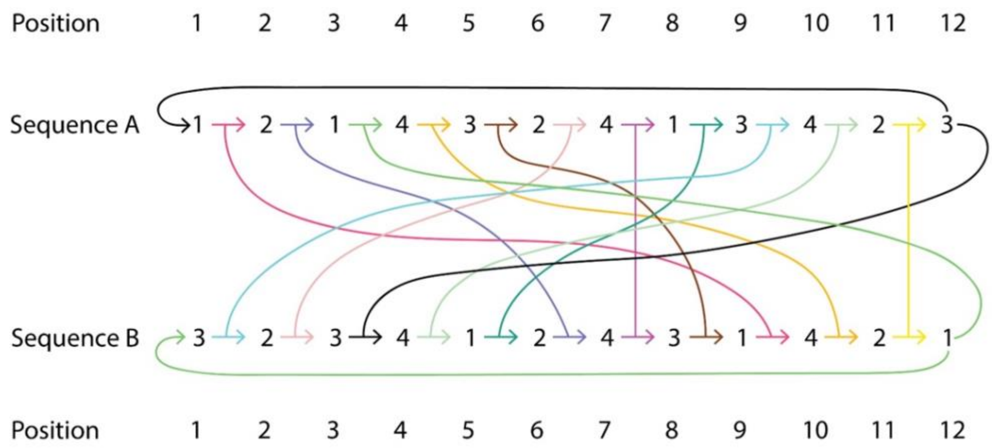

Figure 8. SRT training (A) and control (B) sequence with possible transitions

The task includes instruction slides with video animations to explain the procedure, a practice phase, and a testing phase. The practice phase comprises 60 practice trials to familiarize participants with the task mechanics. The trials follow the training condition 50% of the time, and control condition 50% of the time. The testing phase consists of 8 blocks, each comprising 120 trials, for a total of 960 trials. A short break is given between blocks. The SRT task is scored by subtracting the mean reaction times (RTs) in the training condition from those in the control condition, which indicates the amount of implicit learning.

3.3.6. Forward digit span task

The Forward Digit Span task was used to measure phonological working memory (see Section 2.3.2.1). The format of the task was adapted from the letter span task in Linck et al. (2013), which itself was based on the operation span task used in Unsworth et al. (2005). The task required participants to first listen to, and then repeat, a series of auditory stimuli.

Participants were presented with a series of number sequences in their L1 (Croatian), varying in length from three to nine digits (see Figure 9). The task consisted of seven sets, with each set containing four sequences of equal length. The sequences gradually increased in length between sets, starting with three digits in the first set and progressing to nine digits in the last set, resulting in a total of 28 sequences. A partial-credit scoring system was used, in which points were awarded for any correctly recalled digits in their respective positions. This scoring method was preferred over an absolute scoring system because it offers greater reliability and better discrimination of participants' performance (Conway et al., 2005). The maximum possible score on the task was 168.

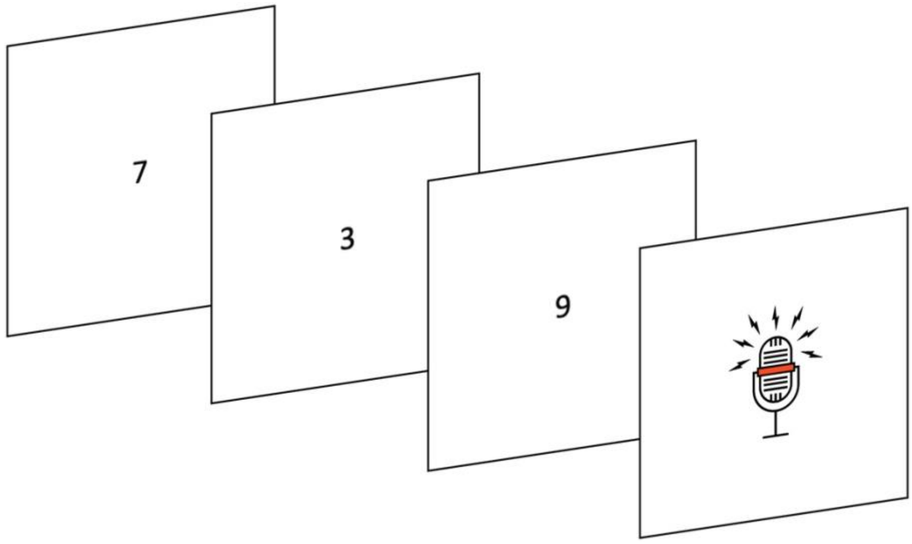

Figure 9. Forward digit span task interface

3.3.7. Operation span task

The Operation Span task was used to measure executive working memory (see Section 2.3.2.2). The task followed the format of Unsworth et al. (2005), which is an automated version of the original word span task proposed by Turner and Engle (1989). Participants were required to solve a series of math operations while simultaneously remembering a set of letters. Unlike the original task, which involved recalling words (Turner & Engle, 1989), this task used letters, as previous research indicated that word knowledge could influence results (Engle et al., 1990). This change helps to avoid conflating long-term memory with executive working memory.

The task began with a practice phase comprising three stages:

1. Math Operations Practice: Participants solved math problems (e.g., (1 * 2) + 1 = ?) and, on the next screen, indicated whether the proposed solution was correct or incorrect. Audio feedback was provided for each response.
2. Simple Letter Span Practice: Letters were presented sequentially on the screen, each remaining visible for 800 milliseconds (following Unsworth et al., 2005). Participants were required to recall the letters in the order in which they appeared. During the recall phase, a 4x3 matrix of letters (F, H, J, K, L, N, P, Q, R, S, T, Y) was displayed, and

participants indicated which letters appeared and in what order. Feedback was provided after each response.

3. Combined Task Practice: Participants practiced both math operations and letter recall together (see Figure 10).

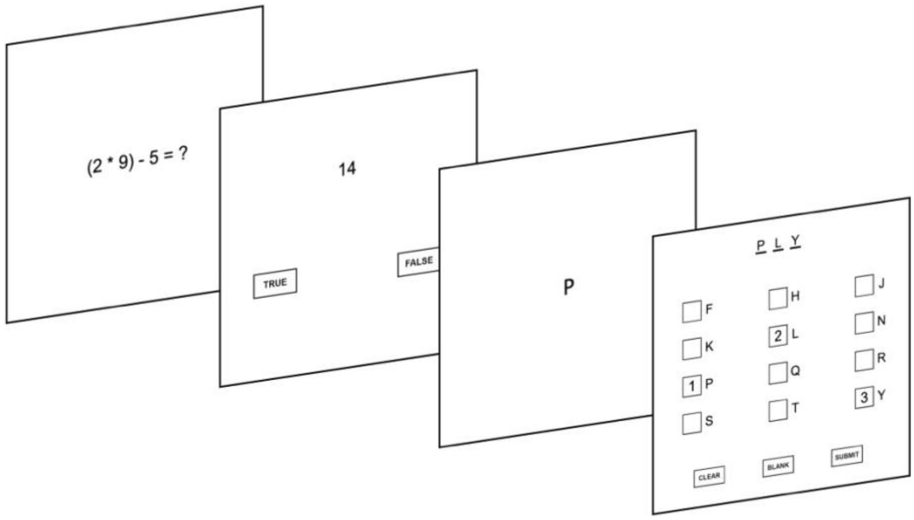

Figure 10. Operation span task interface

During the test phase, participants were presented with a math operation and a proposed solution, after which they had to indicate whether the solution was correct. They were then presented with a letter to remember for later recall. The test consisted of 6 sets, each containing four sequences of the same length. The sequence length started with three letters and gradually increased to eight letters. In total, there were 99 letters and an equal number of math operations to solve. To prevent cognitive overburden in participants with lower executive working memory capacity, the task was automatically aborted if a participant failed to recall any letters correctly in two consecutive sequences.

An accuracy criterion of 85% correct math operations was set to ensure proper engagement in the task. No participants were excluded based on this criterion. For each correctly recalled letter in its proper position, participants earned one point, with a maximum possible score of 99 points.

3.3.8. Oral production task

An oral production task was used to assess participants' overall speaking proficiency. This task required participants to engage in a monologic speaking exercise, in which they were given three minutes to speak uninterruptedly about their typical day at school. Their speech was audio-recorded and later transcribed. The transcripts were analysed using the CLAN software (MacWhinney, 2000), and measures of complexity, accuracy, and fluency (CAF) were calculated as follows:

- Lexical complexity was calculated using Guiraud's index (Guiraud, 1954).
- Morphosyntactic complexity was measured based on the number of clauses per c-unit (Foster et al., 2000).
- Lexical and morphosyntactic accuracy were calculated as the number of lexical and morphosyntactic errors per 100 words (adapted from Michel et al., 2007), respectively, and the values were reversed, i.e., 100 - N(errors) so that higher score reflects higher proficiency.
- Fluency was assessed using pruned speech rate, which is the number of spoken words per minute (adapted from Freed, 2000).

3.3.9. Oxford Placement Test

The Oxford Placement Test (Oxford University Press, n.d.) is designed to assess students' English language proficiency. It consists of two parts: (1) Use of English, which focuses on assessing reading comprehension, and (2) Listening, which measures participants' listening comprehension skills. The reading comprehension section evaluates participants' understanding of grammatical form and meaning, implied meaning, and overall comprehension of the text. The listening comprehension section includes questions based on 10 long dialogues and 5 short monologues, probing participants' understanding of the content and the overall message being communicated.

Both sections aim to assess how well students grasp the meaning of the spoken and written language, making the test an effective indicator of general language ability (Oxford University Press, n.d.). Additionally, the test is adaptive, meaning that the difficulty of questions adjusts dynamically based on the student's responses. The final score comprises the combined results of both parts, as well as separate scores for each section. The maximum score for each section is 120.

3.3.10. Background questionnaire

The background questionnaire collected demographic information such as participants' age and the duration of their English language study. Additional questions inquired about other languages participants may have studied besides English and the length of any stays in English-speaking countries. The demographic information arising from the questionnaire was reported in Section 2.1. above.

3.3.11. Ethical considerations

Prior to completing a background questionnaire, participants received a participant information sheet outlining the purpose of the study and seeking their consent. They were informed about the voluntary nature of their participation and reassured that they could withdraw at any point without providing an explanation. For participants under the age of 18, parental consent was obtained through a written consent form. Ethical approval was granted by the University of Essex, Social Sciences Ethics Sub-Committee for the pilot study (reference: ETH2021-0304) and for the main study (reference: ETH2122-0014). Participants provided informed written consent.

3.4. Pilot study

A pilot study was conducted with a separate sample of 22 Croatian students of English that matched participants from the main study on all accounts prior to the main study to evaluate all the measures, ensure their validity, and test their levels of reliability. The participant profile matched that of the main study: participants ranged in age from 15 to 18 years (M = 15.33, SD = 1.15) and had been studying English for between 8 and 12 years (M = 9.38, SD = 1.24). Additionally, scores on the language measures were analysed to confirm that the difficulty level was appropriate and that the items functioned as intended. Compared to the procedure used in the main data collection, the pilot study did not include an oral production task, and phonological working memory was assessed using a letter span task (Linck et al., 2013) instead of the forward digit span task described in Section 3.3.6. The only difference between these tasks is that the letter span task used letters instead of numbers. Executive working memory was measured with the backward digit span task (Kormos & Sáfár, 2008), rather than the operation span task used in the main study. In this task, participants listened to a sequence of numbers and repeated them in reverse order, thereby assessing both memory span and central executive processes associated with working memory. The remaining measures were identical to those described above, and the procedure followed steps shown in Figure 11.

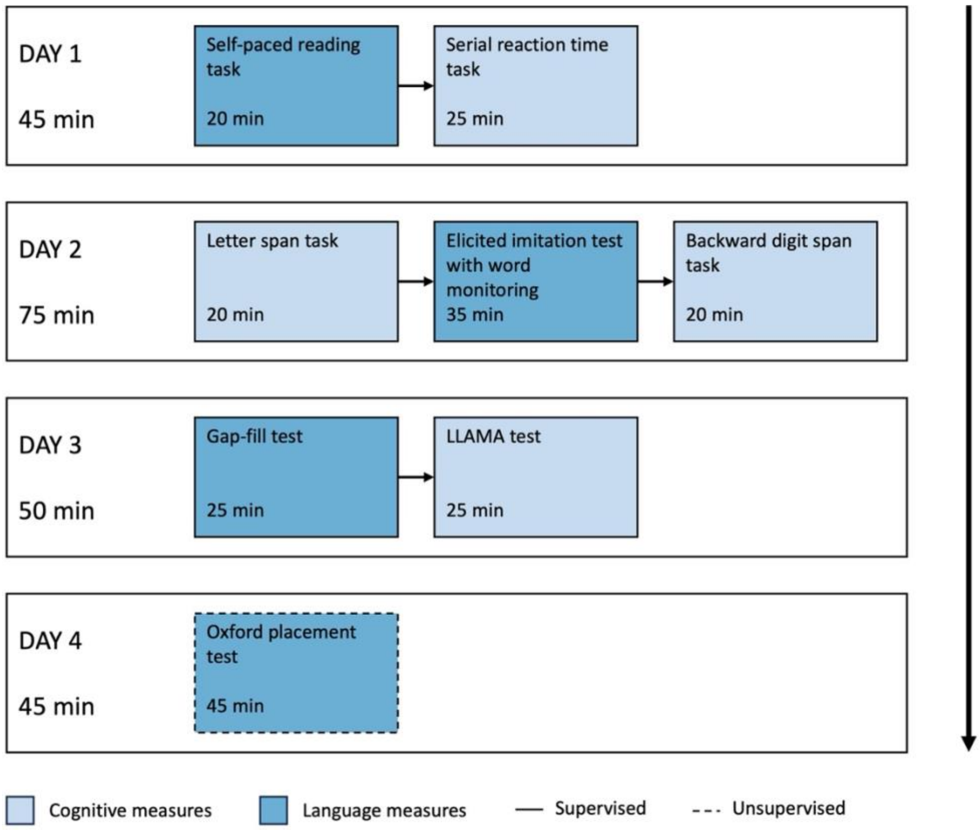

Figure 11. Pilot study procedure

3.4.1. Results

3.4.1.1. Self-paced reading task

Self-paced reading scores were calculated as the difference in reaction time (RT) between grammatical and ungrammatical sentences, yielding a grammatical sensitivity index. Internal consistency (Cronbach's alpha) was high for both list 1 ($\alpha = .996$) and list 2 ($\alpha = .993$), with a total of 96 items (k = 96). A paired t-test was conducted to determine whether the difference between the two conditions was statistically significant, revealing a positive GSI of 15.6 with a very small effect size, $t(21) = 3.112$, $p < 0.005$, Cohen's $d = 0.07$. The positive GSI for the three structures indicates that participants experienced delays when reading grammatically incorrect sentences compared to their correct counterparts.

To further explore this finding, a series of paired-samples t-tests were conducted to examine the RT differences between grammatical and ungrammatical items at different regions of interest (ROIs) and for different structures. The results are presented in Table 6.

Table 6. Descriptive statistics and RT comparison for grammatical and ungrammatical sentences in the SPR task, divided by linguistic structure and ROI

		GRAMMATICAL		UNGRAMMATICAL						
		M	SD	M	SD	GSI	n	t	Sig.	S-W
Articles	ROI 1	519.8	229.3	512.1	235.8	-7.7	22	.628	.537	.012
	ROI 2	486.7	175.2	495.0	191.3	8.4	22	-.931	.363	
	ROI 3	511.2	217.6	544.2	255.3	32.9	22	-2.059	.052	
	ROI 4	503.9	189.8	525.0	197.8	21.1	22	-1.384	.181	
Past	ROI 1	503.9	218.1	514.5	212.6	10.6	22	-.766	.452	.011
	ROI 2	465.6	182.1	490.4	203.5	24.8	22	-1.846	.079	
	ROI 3	509.5	200.3	531.1	255.2	21.5	22	-1.173	.254	
	ROI 4	486.2	184.2	517.7	210.2	31.5	22	-2.396	.026*	
Passive	ROI 1	498.6	240.4	486.4	195.1	-12.2	22	.761	.455	.161
	ROI 2	473.5	181.9	483.7	215.3	10.2	22	-.841	.41	
	ROI 3	503.3	216.6	504.5	234.6	1.2	22	-.093	.927	
	ROI 4	490.7	194.8	535.7	200.4	45.0	22	-3.524	.002**	
Combined		496.1	202.5	511.7	217.3	15.6	22	-3.112	.005**	.010

Note. GSI – Grammaticality Sensitivity Index; S-W – Shapiro-Wilk normality coefficient

As shown, no reliable differences were observed in participants' RTs at the first two ROIs across the structures, which is consistent with previous studies employing SPR (Jiang, 2007; Vafaee et al., 2017). For articles, a marginally significant difference was found at ROI 3, $t(21) = 2.059$, $p < 0.052$, Cohen's $d = 0.14$, with a GSI of 32.9. In the past tense structure, a statistically significant difference emerged at ROI 4, $t(21) = 2.396$, $p < 0.026$, Cohen's $d = 0.16$, with a GSI of 31.5. For the passive structure, a statistically significant difference was also found at ROI 4, $t(21) = 3.524$, $p < 0.002$, Cohen's $d = 0.23$, with a GSI of 45.0. These results suggest that participants were sensitive to grammatical violations for all three structures under investigation.

3.4.1.2. Word monitoring task

Scores on the word monitoring task were calculated as a grammatical sensitivity index, derived by subtracting the mean reaction time (RT) for grammatical items from the mean RT for ungrammatical items (Granena, 2013b; Suzuki & DeKeyser, 2015). Internal consistency (Cronbach's alpha) for the article, past, and passive structures in List 1 (n = 24) was $\alpha = .797$, $\alpha = .672$, and $\alpha = .721$, respectively. For List 2 (n = 24), the values were $\alpha = .171$, $\alpha = .641$, and $\alpha = .645$, respectively. This indicates that the reliability of the task was acceptable across indices, except for the articles in List 2, which was attributed to specific items used in the list.

A paired-samples t-test was conducted to determine whether there was a statistically significant difference in reaction times between the two conditions – grammatical and ungrammatical sentences (Table 7).

Table 7. Descriptive statistics and RT comparison for grammatical and ungrammatical sentences in the word-monitoring component of the EIM test

	GRAMMATICAL		UNGRAMMATICAL						
	M	SD	M	SD	GSI	n	t	Sig.	S-W
Articles	985.4	92.6	1176.6	59.0	191.2	16	-6.727	.000**	.111
Past	1103.8	88.2	1133.3	100.4	29.5	16	-1.416	.177	.856

Passive	1083.1	64.5	972.9	113.0	-110.2	16	4.273	.001**	.001
Combined	1057.4	62	1094.3	63.3	36.9	16	-3.165	.006**	.196

As shown, a statistically significant difference was observed for the three structures combined, with a medium effect size (Cohen's *d* = 0.589) and a GSI of 36.9. When analysing each structure separately, a significant difference with a large effect size was found for articles (*d* = 2.463) and a GSI of 191.2, while the passive voice showed a smaller but still large effect size (*d* = 1.198) with a GSI of -110.2. However, the past simple did not yield a statistically significant difference, and the GSI was 29.5.

Notably, the passive structure displayed a reversed effect where participants were faster on ungrammatical sentences than on grammatical ones, indicating no processing delays for the ungrammatical sentences. This finding contrasts with the results from the self-paced reading task, where delays in RTs for the passive structures were observed, but only three words after the grammatical error (ROI 4). Since the word monitoring task logs RTs at the word immediately following the error, it may not capture the spillover effects that were recorded in the self-paced reading task. Thus, while the word monitoring task effectively registers delays in detecting grammatical errors for shorter structures (e.g., articles), it may not be the best choice for lengthier structures (e.g., passive voice) as it cannot capture any spillover effects.

3.4.1.3. Elicited imitation test

Elicited imitation (EI) scores were computed following the conventions described in Section 2.3.2. above. Internal consistency, as measured by Cronbach's alpha, was α = .804, α = .684, and α = .721 for articles, past, and passive structures, respectively, in List 1. For List 2, alpha values were somewhat lower at α = .694, α = .425, and α = .719 (n = 24), respectively. The mean score was 30.56 (SD = 3.47) for grammatical items, and 24.11 (SD = 5.20) for ungrammatical items. The positive difference between the mean scores for grammatical and ungrammatical items (GSI = 6.45) indicates that participants were more successful at reconstructing sentences when the original sentence was grammatically correct, a phenomenon frequently observed in studies employing elicited imitation (Erlam, 2006; Suzuki & DeKeyser, 2015).

A series of paired-samples t-tests was conducted to determine whether the differences between grammatical and ungrammatical sentences were statistically significant for each structure and across all structures combined. The results, presented in Table 8, show that for all three structures, the differences between grammatical and ungrammatical items were statistically significant. This finding supports the view that elicited imitation is a measure of reconstructive processing, suggesting that the task taps into the participants' L2 system rather than allowing mere rote repetition or reliance on working memory (Erlam, 2006; Suzuki & DeKeyser, 2015).

Table 8. Descriptive statistics and accuracy comparison for grammatical and ungrammatical sentence in the elicited imitation component of the EIM test

	GRAMMATICAL		UNGRAMMATICAL						
	M	SD	M	SD	GSI	n	t	Sig.	S-W
Articles	8.89	2.17	6.56	2.43	2.330	18	4.351	.001**	.179
Past	10.83	0.92	9.17	2.04	1.660	18	3.644	.002**	.370
Passive	10.83	1.04	8.39	2.44	2.440	18	3.661	.002**	.100
Combined	10.19	1.16	8.04	1.73	2.150	18	7.558	.001**	.312

To confirm that the task indeed measured the intended linguistic knowledge through reconstructive processing, two further criteria were applied based on Suzuki and DeKeyser (2015): (1) a high positive correlation between grammatical and ungrammatical sentences (Erlam, 2006), and (2) the absence of a correlation between EI scores and working memory capacity (Okura & Lonsdale, 2012).

Correlation analysis revealed a strong positive correlation between grammatical and ungrammatical sentences, $r = 0.72$, $p < 0.001$, $n = 18$. This indicates that participants who performed well on grammatical items also tended to perform well on reconstructing ungrammatical sentences. Additional correlation analyses found no statistically significant relationship between EI scores and working memory scores, suggesting that success on the EI task was independent of working memory capacity, as expected.

In summary, these findings confirm that the elicited imitation task, as operationalized in the pilot study, measured reconstructive processing ability and functioned as intended.

3.4.1.4. Gap-fill test

Internal consistency for the gap-fill test was acceptable for each structure and for the combined scores across structures. The distribution of scores for the past and passive structures showed a negative skew, indicating that participants tended to perform on the higher end of the scale. A similar pattern, though to a lesser extent, was observed for articles and for the combined scores. The overall mean accuracy score on the gap-fill test was 20.93 out of 25 points, or 83.7%, across the three structures. Descriptive statistics for each structure are presented in Table 9.

Table 9. Descriptive statistics of the gap-fill test results

	n	k	M		SD		Skew	S-W (p)	α
			Raw	Corr.	Raw	Corr.			
Articles	20	25	20.05	80.2	3.4	13.6	-0.631	.079	.760
Past	20	25	22.60	90.4	2.8	11.3	-1.741	.001	.767
Passive	20	25	20.15	80.6	3.7	14.9	-0.711	.033	.783
Combined	20	75	20.93	83.7	3.5	14.0	-0.572	.221	.885

Participants were most accurate on the simple past items, with an average score of 22.6 out of 25, while they scored slightly lower on passive voice (M = 20.15) and articles (M =

20.05). Given the non-normal distribution of gap-fill scores for two out of the three structures, a Kruskal-Wallis test was conducted to determine whether there were significant differences in scores and whether articles were indeed more challenging than past and passive. The differences between the rank totals of 25.23 (articles), 39.53 (past), and 26.75 (passive) were significant, $H(2, n = 60) = 8.228$, $p = .016$. Post hoc comparisons using Mann-Whitney tests with a Bonferroni correction revealed that the difference between articles and past was statistically significant ($z = -3.389$, $p = .002$), as was the difference between articles and passive ($z = -2.788$, $p = .016$). No significant difference was found between past and passive scores. These results indicate that articles were the most difficult structure, followed by passive voice, with simple past being the easiest.

Regarding subjective measures of awareness, participants in this study were generally confident in their responses, with an average confidence rating of 88% across the three structures. The differences in confidence ratings between structures were minimal, as shown in Table 10.

Table 10.Source attributions and certainty ratings on the gap-fill test (pilot study)

	Source attribution (%)				Certainty
	Rule	Memory	Guess	Intuition	rating
Articles	55.0	20.2	4.8	20.0	87
Past	57.8	17.6	4.6	20.0	89
Passive	59.6	18.4	2.4	19.6	88
Combined	57.5	18.7	3.9	19.9	88

As seen in the Table 10, participants primarily relied on their knowledge of rules, followed by intuition and memory, while only 4% of their responses were based on guessing. To further investigate the relationship between GAP scores, confidence ratings, and source attributions, bivariate correlations were calculated between them for the combined structures (Figure 12).

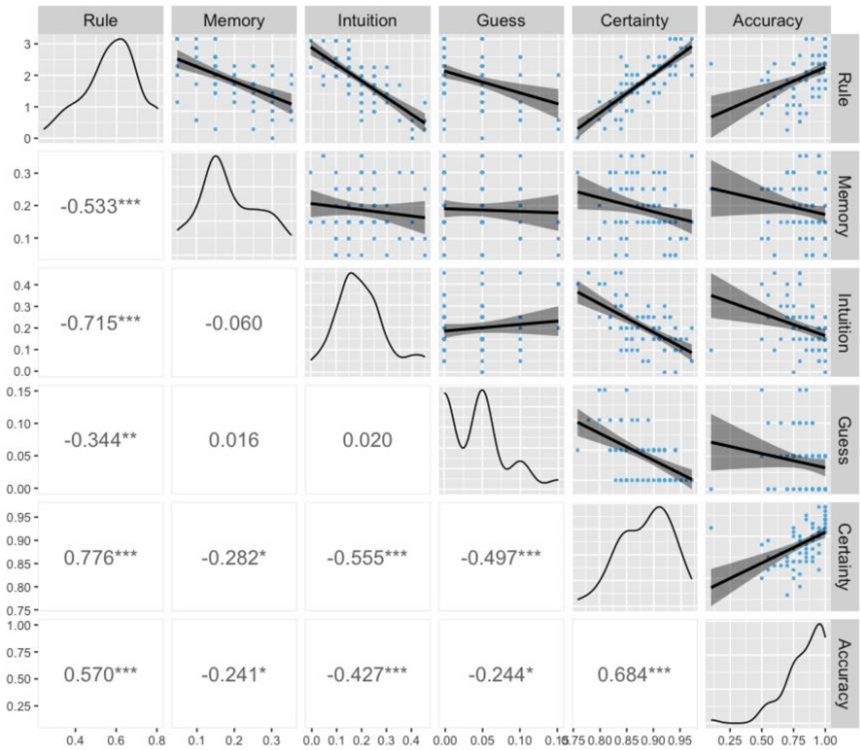

Figure 12. Correlations (Spearman's rho) between source attributions, certainty ratings, and accuracy of gap-fill scores (pilot study)

The results show that high confidence ratings were significantly and positively correlated with gap-fill scores, both for the combined structures and for each structure individually. This indicates that participants who were more certain of their answers tended to be more accurate, suggesting an awareness of their own knowledge. Analysis of the relationship between gap-fill scores and source attribution revealed a similar pattern: reliance on rules was positively and significantly correlated with higher scores, whereas guessing, intuition, and memory were significantly negatively correlated with scores. This suggests that reliance on explicit knowledge of rules was the primary factor contributing to success on the gap-fill test. The significant positive relationship between gap-fill scores and participants' confidence in their answers further supports the notion that the gap-fill test measures explicit knowledge.

3.4.1.5. Serial reaction time task

Scores on the SRT test were calculated separately for each participant across the eight blocks. The mean reaction time for the training condition was M = 453 ms, SD = 19, while the mean reaction time for the control condition was M = 460 ms, SD = 19. An overview of the reaction time data across blocks is presented in Figure 13.

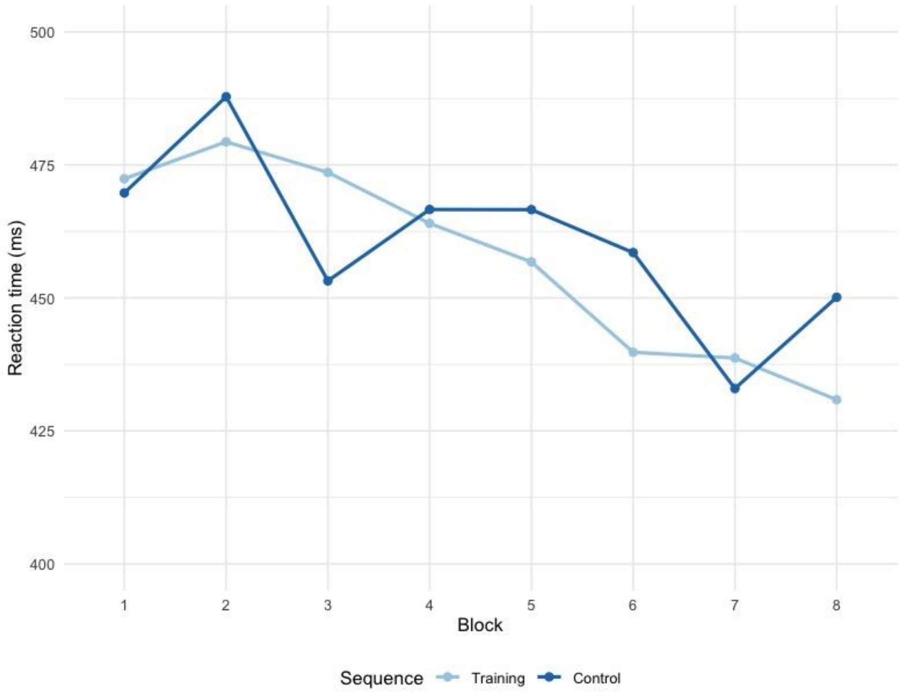

Figure 13. Mean RTs for training and control sequences on the SRT task (pilot study)

Split-half reliability with Spearman-Brown correction yielded a coefficient of α = .245 (n = 8), which is relatively low, even compared to the typically low reliability scores reported for the SRT in other studies (Kaufman et al., 2010; Suzuki & DeKeyser, 2015, 2017), where scores usually hover just above α = .4. This low reliability is likely due to the small sample size (n = 22).

A series of paired-samples t-tests were conducted to examine potential differences between the training and control conditions across the eight blocks. As shown in Table 11, statistically significant differences were observed in blocks 2, 3, 5, 6, and 8, with small effect sizes in all blocks except for block 8, which showed a medium effect size.

Table 11. RT comparison for training and control sequences by block in the SRT task

		M	SD	t	df	Sig.	d
B1 - (85)	B1 - (15)	0.83	25.49	0.150	20	0.882	0.014
B2 - (85)	B2 - (15)	-17.32	30.93	2.565	20	0.018*	0.271
B3 - (85)	B3 - (15)	15.86	22.95	3.167	20	0.005**	0.254
B4 - (85)	B4 - (15)	-5.03	33.37	0.690	20	0.498	0.089
B5 - (85)	B5 - (15)	-14.39	28.11	2.346	20	0.029*	0.229
B6 - (85)	B6 - (15)	-21.02	22.19	4.341	20	0.001**	0.385
B7 - (85)	B7 - (15)	5.01	22.71	1.010	20	0.325	0.097
B8 - (85)	B8 - (15)	-23.02	28.09	3.754	20	0.001**	0.539

The results suggest that the effect size in blocks where a statistically significant difference was found tended to increase as the task progressed. This trend was only disrupted in blocks 3 and 7, where, unexpectedly, the mean reaction time in the control condition was lower than in the training condition. Apart from these exceptions, the increasing difference in reaction times between training and control conditions as the task advanced indicates some degree of automatization. Based on the changing patterns in reaction times between block 1 and block 3, it was concluded that the learning effect became evident from block 4 onwards.

3.4.1.6. LLAMA battery

The LLAMA test was based on version 2. Participants scored the highest on LLAMA B, which measures associative memory, while the lowest scores were observed on LLAMA E, assessing phonemic coding ability. Reliability indices (Cronbach's alpha) were below the acceptable level (<.7) for all subtests except LLAMA B, which is consistent with previous research (see

Bokander & Bylund 2020 for in-depth reliability analysis) utilizing version 2 of the LLAMA test. Detailed descriptive statistics are provided in Table 12.

Table 12. Descriptive statistics of scores across LLAMA battery subtests

	n	k	M		SD		Skew	S-W	α
			Raw	Corr.	Raw	Corr.		(p)	
LLAMA B	21	20	10.10	50.48	4.64	23.18	-0.436	.111	.807
LLAMA D	21	30	21.62	44.13	3.44	20.94	0.472	.200	.552
LLAMA E	21	20	11.90	20.24	2.91	22.95	0.955	.006	.495
LLAMA F	21	20	14.14	38.57	2.87	24.30	-0.617	.003	.582

3.4.1.7. Letter span and backward digit span task

Scores on the working memory measures indicated that participants were highly successful in completing the tasks, with overall high scores. Table 13 presents descriptive statistics, normality analysis, and reliability indices for these measures.

Table 13. Descriptive statistics of scores on letter span and backward digit span tasks

	n	k	M		SD		Skew	S-W	α
			Raw	Corr.	Raw	Corr.		(p)	
LST	14	28	112.57	67.00	20.82	12.40	0.613	.189	.834
BDST	16	28	117.88	70.16	24.41	14.53	0.387	.107	.854

3.4.1.8. Oxford Placement Test

The mean overall English language proficiency score was 184.68 (SD = 26), corresponding to the C1 level on the CEFR (Common European Framework of Reference for Languages). For the Use of English component, the mean score was 100.11 (SD = 16.48), and for Listening, the mean score was 84.58 (SD = 12.13), with a maximum possible score of 120 for both categories. Of the participants, 21% scored at the B2 level, 53% at C1, and 26% at C2 proficiency level.

Overall, participants' scores were on the higher end of the proficiency spectrum, which was contrary to our expectations. This could be attributed to the fact that the participants had been studying English for an average of M = 9.43 years, and since the OPT primarily assesses receptive skills, these scores are indicative of strong passive language abilities.

3.4.2. Concluding remarks

The purpose of the pilot study was to assess the validity and reliability of all measures before the main data collection. Both stimulus lists in the self-paced reading task achieved an excellent reliability index of $\alpha = .9$, while the gap-fill test demonstrated acceptable reliability, slightly over $\alpha = .7$. The serial reaction time task showed reliability comparable to other studies, and a series of t-tests comparing reaction times in the two conditions revealed significant differences in most blocks, indicating a desirable implicit learning effect. Both working memory measures showed good reliability indices, around $\alpha = .8$.

At the same time, several shortcomings were identified, leading to some changes being implemented. First, the GSI for the passive voice in the word monitoring task was negative, meaning participants were faster for ungrammatical sentences, contrary to expectations. Notably, the GSI detected in the self-paced reading task for the passive voice was significant only three words after the grammatical violation, whereas in the word monitoring task, it was calculated for the word immediately following the violation. This led to the conclusion that the word monitoring task might be unable to capture delayed hesitation, or the spillover effect, observed with the passive voice in the self-paced reading task. Despite this, the task and stimuli were retained, as the analysis was conducted on a relatively small sample size, where statistical anomalies are more common.

Next, it was decided that the elicited imitation task with word monitoring would only include stimuli from List 1 due to the relatively low reliability observed for the articles in the word monitoring task and for the past structure in the elicited imitation task. Although the word monitoring task was retained in the main study, more than 40% of participants misunderstood the instructions, resulting in unusable data. As a result, scores from this task were excluded from all analyses, making the self-paced reading task the only measure of implicit knowledge in the final dataset.

Additionally, official changes to the LLAMA battery coincided with the pilot study. Specifically, LLAMA version 3 was published just as the analyses were being conducted. Version 3 reportedly offered improved reliability for the subtests (Rogers et al., 2023), prompting the decision to replace LLAMA version 2 with version 3.

Furthermore, during the transcription of the Backward Digit Span Task (BDST) data, participants were observed employing various strategies to reverse and memorize the numbers, such as reversing and repeating triplets or writing down the numbers. As a result, the BDST was replaced with the Operation Span Task, in which mathematical operations are performed between the presentation of stimuli and the response which makes the use of memorization strategies harder. The task instructions were also modified to explicitly prohibit participants from using any external aids, such as pen and paper. Given that the executive working memory measure now used letters, the stimuli in the Letter Span Task were changed from letters to digits.

Finally, due to the high proficiency scores obtained on the Oxford Placement Test, it was determined that an additional measure of L2 proficiency tapping into productive skills was necessary. To ensure the overall test duration remained reasonable (below 180 minutes), a 3-minute oral production measure was introduced, as described in Section 2.3.8. above.

3.5. Main study procedure

Following the methodological changes informed by the pilot study, the main study was conducted from December 2021 to April 2022 with a sample of 86 L1 Croatian learners of L2 English described in Section 3.1 above.

All tasks were initially designed using Python in the PsychoPy software (Peirce et al., 2019) and were later adapted into an online version using the PsychoJS framework. The tests and experiments were hosted and executed on the Pavlovia platform (Peirce et al., 2019). All tests were supervised by the main researcher via Zoom, except for the Oxford Placement Test, which is hosted on Oxford's official website.

Test administration was divided into three one-hour sessions. After participants completed the consent form and background questionnaire, the linguistic tasks were administered in a fixed order, progressing from the most implicit to the more explicit linguistic tasks interspersed with the aptitude, WM, and proficiency measures: the first session included the self-paced reading task and serial reaction time task; the second session comprised the oral production task, elicited imitation with word monitoring task, forward digit span task, and operation span task; and the third session included the gap-fill test and the LLAMA battery (Figure 14).

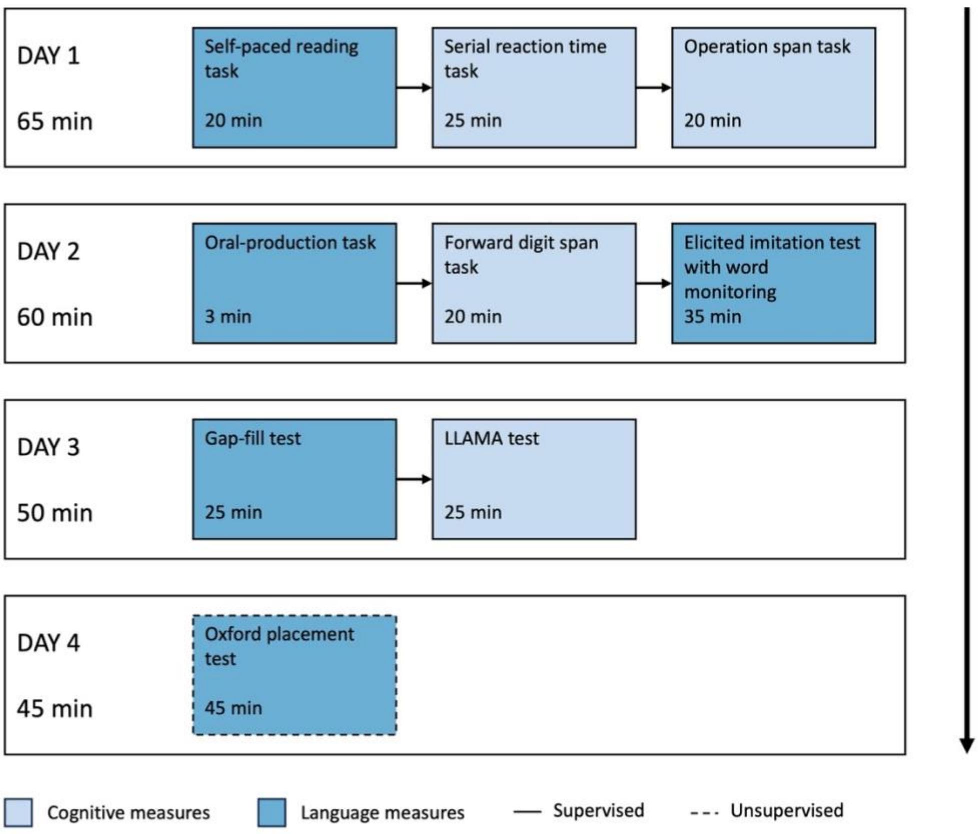

Figure 14. Main study procedure

The Oxford Placement Test was not supervised; a 90-minute time limit was set via the test's settings, and participants were instructed to complete it at their own pace. In total, it took approximately 180 minutes to complete the linguistic and cognitive tests, an additional 10 minutes to fill out the background questionnaire, and around 45 minutes to complete the Oxford Placement Test. As compensation for their participation, students received an official language certificate from Oxford based on their proficiency test score. This was agreed upon in consultation with their English teachers, as it was considered the most beneficial form of compensation.

4. Chapter 4 (Article 1) – Aptitude for explicit and implicit learning

4.1. Abstract

The present study examined the structure of and relationship between aptitude for explicit and implicit learning and working memory. Furthermore, we investigated to what extent these variables could predict second-language (L2) proficiency in terms of reading, listening and grammar knowledge. A total of 86 Croatian learners of English at advanced levels completed the LLAMA aptitude test suite, a probabilistic serial reaction time (SRT) task, operation span and forward digit span tasks, as well as grammar, reading and listening comprehension tests. Our factor-analytic results support a conceptual distinction between (1) working memory, (2) explicit aptitude and (3) implicit aptitude, while at the same time highlighting the multi-componential nature of implicit aptitude, with factor loadings of LLAMA D and SRT pulling in opposite directions. Regression analyses mirror this pattern of results: Whereas components of explicit aptitude, implicit aptitude and working memory significantly predicted L2 proficiency, LLAMA D, SRT and forward digit span emerged as negative predictors. We argue that these findings support a conceptualization of (implicit) aptitude as a cognitive proclivity rather than as a context-independent ability, in line with both current research and previously proposed multi-dimensional and dynamic perspectives of aptitude.

4.2. Introduction

Recent years have seen renewed interest in language learning aptitude research in the context of additional or second-language (L2) learning. On the theoretical side, the role of working memory (WM) as a potential component of aptitude continues to be debated. Moreover, the proposal of distinct aptitudes for explicit and implicit learning is a current focus point that is of immediate relevance to L2 researchers, given the recognition that both explicit and implicit knowledge and learning are implicated in the attainment of L2 proficiency. No study to date has brought together all of the above strands by including measures of aptitude for explicit learning and implicit learning and phonological and executive WM in a single research design scrutinizing the relationships between these variables. Probing their capacity to predict L2 proficiency assessed in terms of learners' grammar knowledge, reading, and listening skill is also required. This is what the present study set out to do.

4.3. Background

4.3.1. Language learning aptitude and WM as predictors of L2 attainment

Language learning aptitude refers to a set of cognitive and perceptual abilities that facilitate fast and easy learning of new languages (Carroll, 1981). The classic model of aptitude (Carroll, 1981) comprises phonetic coding ability, associative memory, and language-analytic ability (Skehan, 1998). While the predictive power of aptitude has been found to be superior to that of factors such as WM (Linck et al., 2013), motivation (Masgoret & Gardner, 2003) or anxiety (Teimouri et al., 2019), meta-analytic research has shown that different components of aptitude differentially predict L2 skills such as listening, reading, speaking and grammar (Li, 2016). In particular, language-analytic ability strongly predicts L2 grammar learning and reading comprehension, and phonetic coding ability has been found to be a good predictor of vocabulary learning and general L2 proficiency (Li, 2022).

The distinction between aptitude for explicit learning and aptitude for implicit learning is relatively recent. The classic aptitude components of phonetic coding ability, language-analytic ability and associative memory are seen as representing aptitude for explicit learning (henceforth: explicit aptitude). Conversely, aptitude for implicit learning (henceforth: implicit aptitude) refers to cognitive abilities that facilitate implicit L2 processing and learning in the absence of conscious awareness (Granena, 2020). Whereas explicit aptitude may primarily predict achievement at beginner levels (Linck et al., 2013; Robinson, 2005), implicit aptitude is expected to predict ultimate attainment (Li & DeKeyser, 2021). Indeed, two hypothesized measures of implicit aptitude, the serial rection time (SRT) task and LLAMA D (further discussed below), have been found to significantly predict grades achieved in foreign-language classes (Kaufman et al., 2010), speech fluency (Granena, 2019), and pronunciation accuracy (Saito et al., 2019) at intermediate to advanced L2 levels. Therefore, in addition to considering different L2 skills, it is important to take learners' L2 proficiency level into account when interpreting results pertaining to the predictive power of (components of) explicit and implicit aptitude.

WM refers to the ability to simultaneously store and process information while engaging in a cognitive task (Baddeley, 1986; Cowan, 2005). In L2 research, Baddeley's (2015, 2017) multiple-component model of WM has been most influential, with two components of central interest: phonological working memory (PWM), which is responsible for the short-term storage of phonological information and articulatory rehearsal, and executive working memory (EWM), which controls processes such as inhibition, updating and switching (Wen, 2019).

The importance of PWM and EWM in L2 processing, learning and use is well-documented (Juffs & Harrington, 2011; Linck et al., 2014), and the role of WM as an individual-difference variable that can potentially predict L2 outcomes has been acknowledged in aptitude research too (Skehan, 2002, 2016). Furthermore, Robinson's (2005, 2012) model of aptitude complexes argues that different aptitude complexes are dependent on specific combinations of underlying primary cognitive abilities, including WM.

Empirical investigations into the relationship between WM and aptitude have led to mixed results. Several studies have identified no or weak relationships between measures of the two constructs (Roehr & Gánem-Gutiérrez, 2009b; Yoshimura, 2001); Li's (2016) meta-analysis identified a weak correlation between PWM and EWM and overall aptitude and aptitude components.

Yalçın and Spada (2016) found a relationship between EWM measured by first-language (L1) and L2 reading span tasks and language-analytic ability, while no correlation was found between EWM measured by an operation span task and any aptitude component. A study by Sáfár and Kormos (2008) replicated these findings, but did not find a relationship between aptitude and PWM operationalized as a non-word repetition task, a result that was subsequently confirmed (Hummel, 2009). Those of the previously mentioned studies using factor analyses as well as Granena (2013a) found that PWM and EWM loaded on the same factor and separately from overall aptitude or aptitude components. This contrasts with findings reported by Li (2013), where EWM operationalized as a listening span task loaded on the same factor as language-analytic ability.

Taken together, these findings present a mixed picture, no doubt at least partly due to the range of measures used, but also due to differences in participants' profiles, not least in terms of language background and L2 proficiency level. Specifically, PWM appears to be an important predictor of vocabulary, grammar and reading at lower levels of proficiency and/or in novice learners (Hummel, 2009; Serafini & Sanz, 2016), whereas the role of WM at higher levels is less clear. Linck et al. (2013) reported a positive influence of PWM on long-term listening and reading attainment in a group of advanced learners, whereas other studies with experienced learners at advanced levels found no effect of PWM on vocabulary and grammar knowledge (Linck & Weiss, 2011) and no association between EWM and knowledge of a grammatical structure of high learning difficulty (Roehr-Brackin et al., 2021).

Nevertheless, some common threads can be identified. First, (at least some) components of WM are (weakly) related to (at least some) components of aptitude (e.g., Li, 2016; Roehr & Gánem-Gutiérrez, 2009b; Yoshimura, 2001), suggesting a role for WM in L2 learning that is (partly) independent of aptitude – a situation which calls for the inclusion of WM measures in studies aimed at identifying predictors of L2 achievement. Second, including measures of both PWM and EWM seems advisable, given that the two components have been shown to contribute differently to L2 proficiency (Linck et al., 2014). Third, taking into account learners' L2 proficiency level appears to be of critical importance (e.g., Hummel, 2009; Linck et al., 2013; Roehr-Brackin et al., 2021; Serafini & Sanz, 2016).

4.3.2. Measuring aptitude

Studies measuring L2 learners' aptitude have increasingly drawn on the LLAMA battery (Meara, 2005; Meara & Rogers, 2019), a suite of computer-administered tests that is freely available and can be used with participants from a range of L1 backgrounds (Rogers et al., 2017). The LLAMA comprises four subtests that essentially operationalize the classic Carrollian notion of aptitude, that is, associative memory in the sense of vocabulary learning (LLAMA B), phonetic coding ability in the sense of auditory pattern recognition (LLAMA D) and sound-symbol correspondence (LLAMA E) and language-analytic ability in the sense of grammatical inferencing (LLAMA F). LLAMA B, E and F have learning and testing phases, while LLAMA D consists of an exposure and testing phase (with variations in different versions from v.1 to v.3, as discussed below).

While LLAMA B, E and F are regarded as measures of explicit aptitude, it has been suggested that LLAMA D may be a measure of implicit aptitude (Granena, 2013a, 2016), although this view has recently been challenged (Iizuka & DeKeyser, 2023). Another proposed measure of implicit aptitude that seems to be accepted more widely is the probabilistic SRT task (Kaufman et al., 2010), a computer-administered, non-verbal test in which participants react to changes in the location of visual stimuli by pressing keys corresponding to the position of the stimuli on the computer screen. The stimuli follow a probabilistic sequence in an attempt to mirror implicit sequence learning (of language) in the real world (Jiménez & Vázquez, 2005). Unknown to participants, a training sequence is presented 85% of the time, while a control sequence appears for the remaining 15% of the time. Learning is operationalized as faster responses in the training condition compared to the control condition. A growing number of studies employing this measure is testimony to its increasing popularity in L2 research (e.g., Granena, 2013b, 2016, 2019; Linck et al., 2013; Roehr-Brackin et al., 2023; Suzuki & DeKeyser, 2015, 2017; Yi, 2018).

Research to date has reported convergent validity between the SRT task and LLAMA D with measures of implicit knowledge and divergent validity with measures of explicit knowledge (Granena & Yilmaz, 2019; Yilmaz & Granena, 2019), as well as an absence of correlations between the SRT task and tests of explicit aptitude, PWM and EWM (Granena, 2019; Kaufman et al., 2010; Linck et al., 2013; Suzuki & DeKeyser, 2015). The status of LLAMA D in relation to other measures of aptitude and measures of WM is less clear. On the one hand, LLAMA D has been found to be uncorrelated with other LLAMA sub-tests (Saito et al., 2019). On the other hand, it did correlate with PWM, long-term memory retrieval as measured via a semantic priming task, and LLAMA B, while at the same time being uncorrelated with the SRT task (Granena, 2019). Studies drawing on factor analysis likewise show mixed results. LLAMA D loaded on the same factor as a probabilistic SRT task (Granena, 2012; Roehr-Brackin et al., 2023), but on a separate factor than a deterministic SRT task (Granena, 2019).

Taken together, these results could be interpreted as emerging evidence for a multi-componential structure of implicit aptitude (Li & DeKeyser, 2021), with LLAMA D potentially probing implicit memory ability in the verbal domain and the SRT task domain-general implicit learning ability (Granena, 2020). At the same time, seemingly inconsistent findings involving

LLAMA D could be attributable to the test instructions used in any given study and thus be a methodological issue (Li, 2022). Specifically, if participants are informed that they will be tested on the items they hear in the exposure phase, attempts at intentional and therefore explicit learning could ensue. In order to test this hypothesis, Iizuka and DeKeyser (2023) compared three types of LLAMA D instructions ranging from more to less explicit ('listen and memorize', 'just listen', 'sound check') [1] and their effects on task performance. The researchers found that only the 'just listen' instructions that asked participants to carefully listen to the stimuli resulted in a relationship between LLAMA D and the SRT task. However, surprisingly, the relationship was negative. In an attempt to interpret this unexpected result, the researchers suggest that an ability to focus on the auditory stimuli helped with LLAMA D, but had the opposite effect in the case of the SRT task, where focusing on the stimuli on a trial-to-trial basis may have prevented successful (implicit) learning of the probabilistic sequence. Such an interpretation, in turn, might suggest that implicit aptitude is primarily a lack of interference rather than a measurable ability that enhances learning (Iizuka & DeKeyser, 2023). In this regard, the study reports a novel finding and offers an interesting interpretation that could potentially have wide-ranging implications for the conceptualization of implicit aptitude. However, replication is clearly needed.

4.4. The current study

The preceding sections have highlighted several open questions in relation to the theoretical status and empirical measurement of explicit and implicit aptitude. First, the status of LLAMA D and the SRT task as measures of implicit aptitude is still unresolved, leading to the question of exactly how these two tasks relate to each other and, as a consequence, how resulting scores are to be interpreted. Second, the role of WM remains unclear, both in relation to measures of explicit and implicit aptitude and in relation to the relative importance of PWM and EWM at different L2 proficiency levels. Third, the attainment of an understanding of the predictive validity of explicit and implicit aptitude and WM is crucial to the field, yet no study to date has included measures of all these variables in combination with an assessment of several components of L2 proficiency. With a view to addressing these issues, we posed the following research questions:

RQ1: Is there evidence of convergence between auditory pattern recognition ability as measured by LLAMA D and implicit sequence learning ability as measured by a probabilistic SRT task?

RQ2: What is the relationship between measures of aptitude for explicit and implicit learning and measures of WM?

RQ3: To what extent do aptitude for explicit and implicit learning and WM predict L2 proficiency?

4.5. Method

The present study used a correlational design involving the online administration of the LLAMA test suite, a probabilistic SRT task, measures of PWM and EWM, and a measure of L2 proficiency capturing the dimensions of reading comprehension, listening comprehension and morphosyntactic knowledge of selected structures.

4.5.1. Participants

A total of 86 L1 Croatian learners of L2 English participated in the study. At the time of data collection, the participants had been learning English for between 6 and 13 years (M = 10, SD = 1.72) in the context of mandatory classes as a part of their school curriculum. The sample included 62 women, 22 men, and two participants who preferred not to disclose their gender.

Participants were in secondary education and ranged in age from 15 to 18 years (*M* = 16.14, *SD* = 1.29).

4.5.2. Instruments and procedure

All measures with the exception of the L2 reading and listening comprehension tests were programmed into PsychoPy and subsequently administered via the Pavlovia platform (Peirce et al., 2019). All test instructions were provided in L1; the participants were instructed to use headphones in a quiet environment. The first author monitored participants via Zoom to ensure adherence to protocol and allow participants to ask clarification questions. Completion of the L2 reading and listening tests was not monitored because these tests relied on a commercial testing program, as detailed below. Testing proceeded in the following order: SRT task, operation span task (EWM) (Day 1 – c. 50-minutes); forward digit span task (PWM) (Day 2 – c. 20 minutes); gap-fill task (L2 morphosyntactic knowledge), LLAMA (Day 3 – c. 50 minutes); Oxford Placement test (L2 reading and listening) (Day 4 – c. 45 minutes).

4.5.3. Explicit and implicit aptitude

Language learning aptitude was measured by means of the LLAMA suite and a probabilistic SRT task. The LLAMA battery comprises four subtests: LLAMA B, LLAMA D, LLAMA E and LLAMA F.

LLAMA B assesses associative memory, requiring participants to learn 20 new vocabulary items associated with novel picture stimuli during a two-minute learning phase. In the subsequent untimed test phase, participants are presented with a word and must select the corresponding picture from the entire array of 20 pictures. LLAMA B as used in the present study was identical to v.2, except for the removal of the feedback sound in the testing phase. The maximum score was 20, with 1 point awarded for each correct answer and no penalty for guessing.

LLAMA D tests auditory pattern recognition ability. During the exposure phase, participants hear 10 words playing one by one in an unknown language. In the test phase, participants listen to words from the same language, including items heard previously and items not heard before. They respond in a yes/no format to whether an item was familiar or not. Incorrect responses were penalized to compensate for guessing. The feedback sound from v.2 was removed. While this subtest had 30 items in v.1, we included 40 items (i.e., 20 familiar items, each of the 10 items from the exposure phase appearing twice, and 20 unfamiliar items,15 from v.1 and another 5 unused in v.1, but available as downloadable files). The instructions in the present study told participants to listen carefully to the sounds because they would be tested subsequently. The maximum score was 40.

LLAMA E assesses sound-symbol correspondence. Participants are presented with 24 phonetic symbols, each corresponding to a unique syllable. Upon clicking on a symbol, the associated syllable is played. Participants can click on any symbol any number of times during the two-minute learning phase. In the untimed test phase, participants hear a combination of two syllables and must select the correct answer from an array of 20 combinations of previously seen symbols. The version used in the present study was equivalent to v.3. We applied a partial-credit scoring system which awarded one point for each correct syllable in any given two-syllable combination. The maximum score was 40.

LLAMA F is a grammatical inferencing task in which participants have four minutes to work out the rules of an unknown language. During the learning phase, they click on buttons that reveal picture stimuli with corresponding written descriptions. There are 20 items, and participants can click on any button any number of times. During the untimed test phase which comprises 20 items, participants are presented with similar stimuli and must select a combination of words that correctly describes the picture at hand. Participants construct their answers from a board of 16 words. The version used in the present study was equivalent to v.3

except for the fact that all 20 items from v.1 were used. In our partial-credit scoring system, each correct word yielded up to two points: one point for the appropriate word itself, and one point if the word was in the correct position. The maximum score was 132.

The probabilistic SRT task was administered to gauge aptitude for domain-general implicit sequence learning. The task required participants to react to visual stimuli in the form of black squares that appeared in one of four possible locations on the computer screen by pressing a corresponding key as quickly and as accurately as possible. The sequence of stimuli was produced by a probabilistic rule which meant that 85% of the time the stimuli followed a training sequence, while the remaining 15% of the time the stimuli followed a control sequence. Instructions accompanied by video animations and a 60-trial practice phase preceded the task itself, which consisted of 8 blocks, each comprising 120 trials, resulting in a total of 960 trials. There were short breaks between blocks. Following the study protocol from Kaufman et al. (2010), trials were first randomized within their respective block, and subsequently administered in a pre-determined sequence. The task was scored by subtracting the mean response time (RT) in the training condition from that in the control condition.

4.5.4. Working memory

PWM was tested by means of a forward digit span task. We used the format developed by Linck et al. (2013) through adaptation of a component of the operation span task created by Unsworth et al. (2005). Participants were presented with a series of auditory number sequences in L1, varying in length from three to nine digits. There were four sequences of a given length in each set, with the task comprising seven sets, resulting in a total of 28 sequences. Points were awarded for correctly recalled digits in their respective positions. This partial-credit scoring system has been shown to be preferable to an all-or-nothing system due to greater reliability and better discrimination (for details, see Conway et al., 2005). The maximum score was 168.

EWM was assessed by means of an automated operation span task (Unsworth et al., 2005). Participants first solved a simple mathematical problem and then indicated whether the solution shown on screen was correct or incorrect. Subsequently, they were presented with a letter and asked to memorize it. Upon completion of a sequence of mathematical problems followed by letters, participants had to select the memorized letters from an array in the order in which they had been encountered previously. The task comprised 18 sets of letter sequences that ranged in length from three to eight, totalling 99 letters. The maximum score was 99, based on a partial-credit scoring system that awarded points for each correctly recalled letter in a given sequence. Participants' responses to the mathematical problems were used to monitor engagement with the task; a cut-off point of 85% accuracy was set in order to ensure that cognitive resources were duly deployed towards solving the arithmetic equations rather than rehearsing the letters to be recalled.

4.5.5. L2 proficiency

Participants' morphosyntactic knowledge of L2 English was assessed by means of a gap-fill task with a three-way multiple-choice answer format. The test comprised 75 sentences targeting the use of articles, the simple past tense and the passive voice. The choice of targeted structures was based on the participants' grammar syllabus in the context of their English language classes and informed by frequently made mistakes as reported by their teacher (D. Linić Učur, personal communication, June 7, 2020). The maximum score was 75.

The Oxford Placement test (Oxford University Press, n.d.) was used to assess participants' L2 reading and listening comprehension. The reading section draws on test takers' knowledge of grammatical form and meaning, implied meaning and overall reading comprehension. The listening section assesses test takers' listening comprehension through ten dialogues of varying lengths and five short monologues. Both sections are designed to test how well learners understand the meaning of what is being communicated as an indicator of general language ability (Oxford University Press, n.d.). The test is adaptive (i.e., the difficulty of presented items is kept in line with each learner's performance). The test provider's platform generates a total test score as well as separate scores for each section, with a maximum score of 120 for each section.

4.5.6. Data analysis

In order to answer the research questions, reliability indices (Cronbach's alpha) and descriptive statistics were calculated. Normality of data distributions was assessed by means of Shapiro-Wilk tests. Bivariate correlations and exploratory factor analysis were employed to investigate the relationships between variables and to examine the structure of the constructs of explicit and implicit aptitude and WM. Correlations followed by multiple regression analyses were used to establish the predictive power of the aptitude and WM measures with regard to L2 proficiency. The alpha level was set at .05. We conducted the statistical analyses in the R package, v.2021.09.2 (R Core Team, 2021) and IBM SPSS Statistics, v.27.0 (IBM Corp., 2020).

4.6. Results

This section provides descriptive statistics for the cognitive and proficiency measures followed by answers to our three research questions in chronological order.

Participants scored highest on LLAMA B (M = 57.18, SD = 22.34) and lowest on LLAMA D (M = 37.57, SD = 23.10). LLAMA E (α = 0.97) and LLAMA F (α = 0.95) showed excellent reliability; LLAMA B also showed very good reliability (α = 0.81). LLAMA D yielded a lower but still acceptable coefficient (α = 0.72). The full descriptive statistics are shown in Table A in Appendix 2.

Additionally, we calculated participants' SRT task scores and scrutinized differences in mean RT between the eight task blocks to establish the time course of any learning effects. First, error responses (9% of the data) were discarded. Significant outliers (1% of the data), defined as values of more than three SDs from the mean RT for each participant in each block (Granena, 2016; Suzuki & DeKeyser, 2015), were likewise discarded, reducing the sample size to 83. Each participant's SRT score was calculated by subtracting average RT in the training condition from average RT in the control condition. **Error! Reference source not found.** shows mean RTs in each block on the SRT task on both training and control trials.

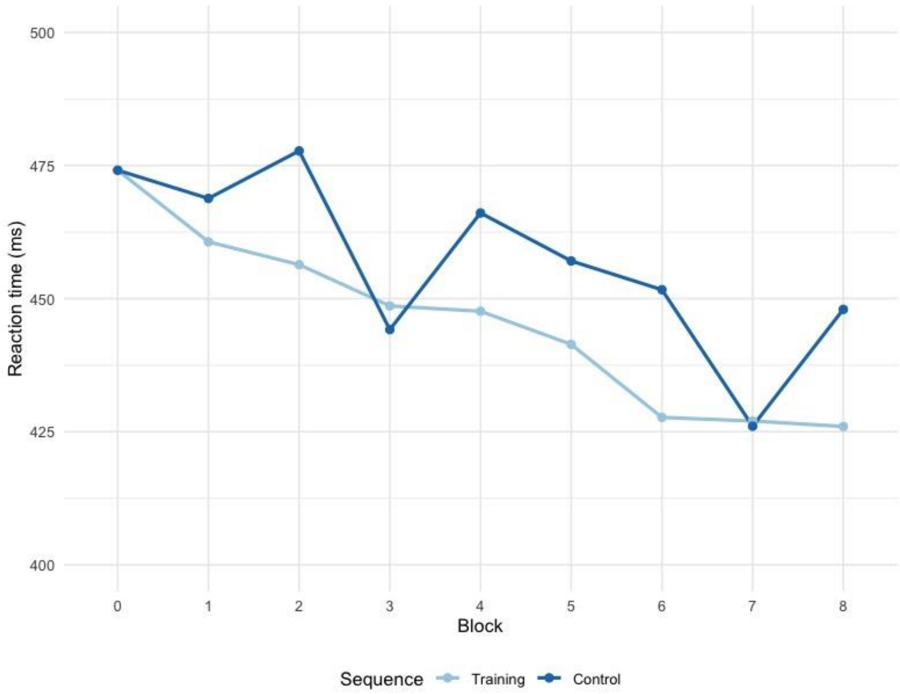

Figure 15. Mean RTs for training and control sequences on the SRT task

As **Error! Reference source not found.** indicates, the RT in the control condition was larger than in the training condition on all blocks except for blocks 3 and 7. The mean RT for the training condition across all blocks was 447ms (*SD* = 67); the mean RT for the control condition was 457ms (*SD* = 71). Table B in Appendix 2 shows mean RTs broken down by block and condition.

Split-half reliability with Spearman-Brown correction resulted in a coefficient of 0.42 for all eight blocks and 0.44 for blocks 4-8. These indices are comparable to the reliability of similar tasks in previous studies (Granena, 2013b; Kaufman et al., 2010; Suzuki & DeKeyser, 2015, 2017). Overall, a reliability coefficient of above 0.4 is considered acceptable for measures of implicit processes, since they typically yield lower indices than measures of explicit processes (Suzuki & DeKeyser, 2015).

A series of paired-samples t-tests was run to identify differences between the training and control conditions in each block. A statistically significant difference was observed in all blocks except blocks 3 and 7. A small effect size was detected in blocks 4 and 5, and a medium effect size in blocks 2, 6, and 8. Cohen's *d* across the last 5 blocks was 0.75, suggesting a medium effect size that is substantially higher than the effect sizes reported in previous research: 0.19 in Kaufman et al. (2010) and 0.21 reported in Suzuki and DeKeyser (2015). As our data show a more stable learning effect from block 4 onwards, all subsequent analyses used scores based on the RT differences from blocks 4 to 8. Table C in Appendix 2 shows the results of the comparisons with effect sizes for each block.

4.6.1. Is there evidence of convergence between auditory pattern recognition ability as measured by LLAMA D and implicit sequence learning ability as measured by a probabilistic SRT task?

Bivariate correlations (Spearman's rho) were run to examine whether there was any convergence between the SRT task and LLAMA D as hypothesized measures of implicit aptitude. Figure 16 shows correlation coefficients (upper triangle), scatterplots for variable pairs (lower triangle) and density plots for each variable (on the diagonal). The results show no significant association between SRT and LLAMA D, thus indicating divergence.

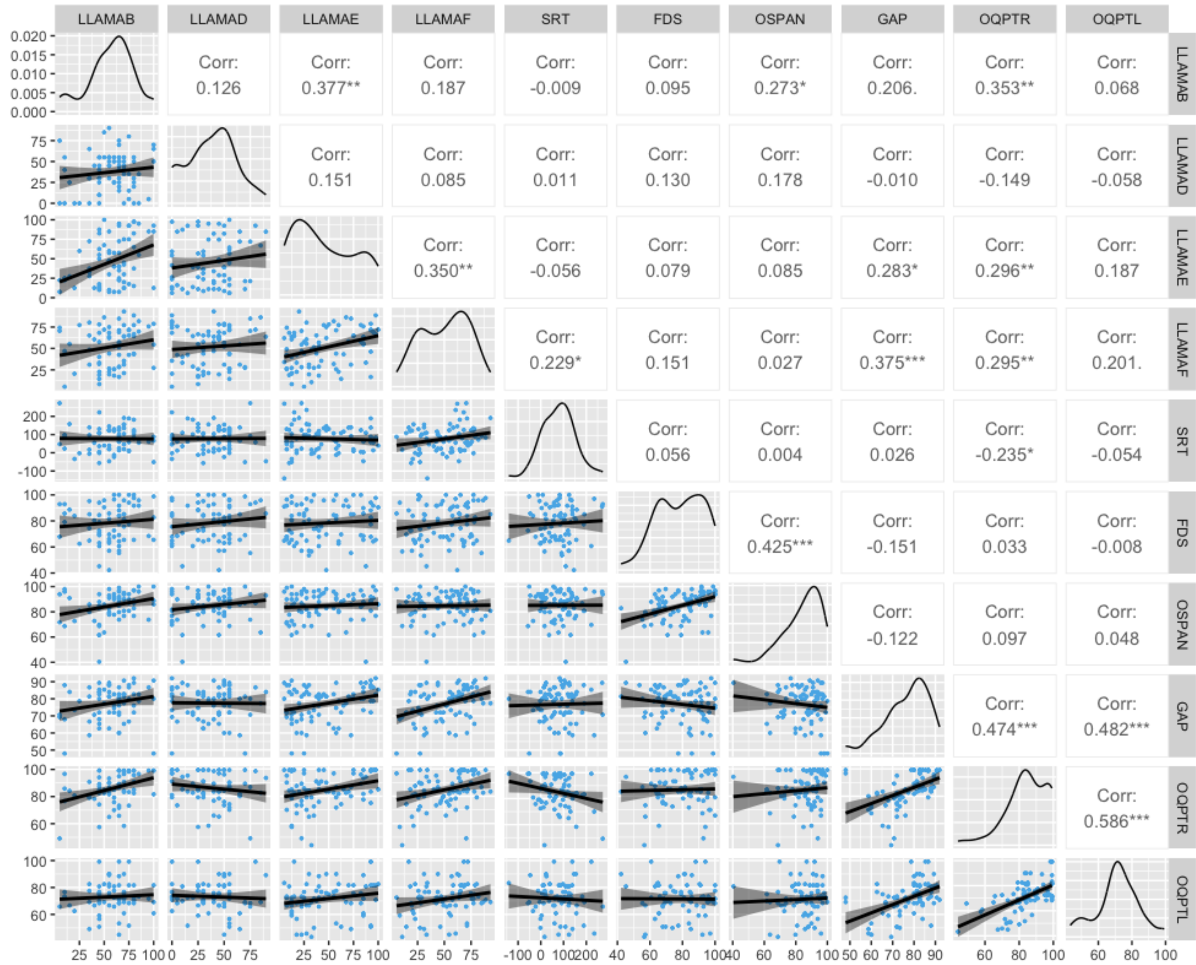

Figure 16. Correlations (Spearman's rho) between measures of aptitude, WM, and L2 proficiency

Finally, we conducted an exploratory factor analysis, using principal component analysis with direct oblimin (oblique) rotation, following confirmation that underlying factors were related (Tabachnick & Fidell, 2014, p. 651). Assumptions were met, with a Kaiser-Meyer-Olkin (KMO) measure of sampling adequacy of 0.55 and Bartlett's test of sphericity significant. The analysis yielded three components with eigenvalues above 1. The factor loadings are shown in Figure 17. A detailed overview of factor loadings can be found in Table D in Appendix 2.

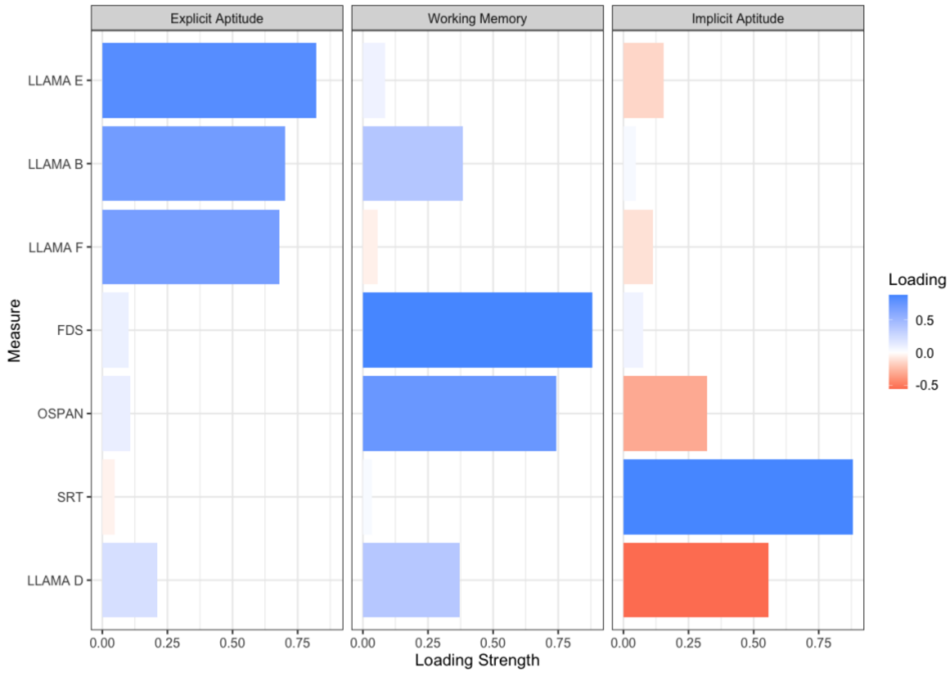

Figure 17. Factor loadings for a three-component solution (PCA)

The SRT task and LLAMA D, the two hypothesized measures of implicit aptitude, loaded on the same factor and distinct from other measures of explicit aptitude (factor 1) and WM (factor 2). While the initial factor loadings were as expected, it is noteworthy that the SRT task and LLAMA D are not correlated (see Figure 16). Furthermore, the factor loading for the SRT task is positive, while the loading for LLAMA D is negative. This discrepancy suggests a complex relationship between the two measures that cannot be fully captured with a simple convergence test.

4.6.2. What is the relationship between measures of aptitude for explicit and implicit learning and measures of WM?

Prior to considering the relationship between the various measures, we calculated descriptive statistics, normality and reliability for the forward digit span (FDS) task that was used to assess PWM and the operation span (OSPAN) task used to assess EWM. Reliability of the FDS task was very good ($\alpha = 0.86$). The OSPAN required the exclusion of three participants who did not meet the 85% accuracy criterion on the mathematical equations that preceded the letter memorization and recall component. Internal consistency was acceptable ($\alpha = 0.73$) and comparable to the reliability indices reported in previous studies: 0.78 in Unsworth et al. (2005) and 0.69 in Suzuki and DeKeyser (2015). Full descriptive statistics can be found in Table E in Appendix 2.

Figure 16 shows the correlations between the measures of aptitude for explicit and implicit learning and the working WM measures. LLAMA B and E, as well as LLAMA E and F, are significantly correlated at a moderate level of strength, which is in accordance with expectations. The two WM measures are likewise positively and significantly associated, as one might expect, and LLAMA B is moderately correlated with the OSPAN, suggesting an association between the ability to learn new lexical items and the central executive component of WM.

As can be seen from the high factor loadings in Figure 17, LLAMA E, B, and F as measures of explicit aptitude load on factor 1 ($\lambda = 1.97$), which accounts for 28% of the variance. The two WM measures load on factor 2 ($\lambda = 1.31$), which explains 19% of the variance. Finally, the SRT task and LLAMA D, conceptualized as measures of implicit aptitude, load on factor 3 ($\lambda = 1.12$), which explains 16% of the variance. Taken together, the three factors explain 63% of the variance and highlight the distinct loading patterns of aptitude measures compared to those assessing WM.

4.6.3. To what extent do aptitude for explicit and implicit learning and WM predict L2 proficiency?

We began addressing the final research question by examining the descriptive statistics for the measures of L2 proficiency used in the present study (gap-fill task, reading and listening sections of the Oxford Placement test), as well as the reliability and normality of the gap-fill task.

The reliability indices for the gap-fill test are all above .98 and therefore deemed excellent. Data were not normally distributed, with a negative skew suggesting a tendency for participants to score at the higher end of the spectrum. With regard to the targeted morphosyntactic structures, articles posed the greatest challenge (M = 61.85, SD = 10.35), while the passive voice was easiest for participants (M = 88.68, SD = 13.27). Scores on the Oxford Placement test were likewise not normally distributed, again due to a negative skew indicative of generally high scores. In terms of the Common European Framework of Reference for language proficiency, 1% of participants were at level B1 ('Threshold'), 16% at B2 ('Vantage'), 45% at C1 ('Effective operational proficiency') and 38% at the highest possible level C2 ('Mastery'). Put differently, 83% of the learners were proficient users of L2 English (i.e., at advanced levels). Full descriptive statistics can be found in Table F in Appendix 2.

Next, we examined the relationships between all variables, as shown in Figure 16. The three L2 proficiency measures are positively associated with each other at a medium level of strength, which is not unexpected. The measures of explicit aptitude LLAMA B, E and F are moderately but significantly correlated with reading, and LLAMA E and F are moderately correlated with the gap-fill test assessing morphosyntactic knowledge. It is worth noting that none of the hypothesized predictor variables is correlated with listening.

Based on the correlation results, statistical (stepwise) regression was deemed unsuitable, as this approach is known to be sensitive to multicollinearity and may fail to identify important predictors or arbitrarily favour one predictor over another (Field, 2018). In contrast, hierarchical regression offers greater control over the order of variable entry and is considered more suitable for replicability. Consequently, we conducted two hierarchical multiple regression analyses with reading and gap-fill as dependent variables, respectively. In each analysis, predictor variables were entered in descending order according to the absolute values of their correlation coefficients with the outcome variable, as shown in Figure 16. All assumptions were met in accordance with Field (2018) and Jeon (2015). Univariate outliers (cases with a z-score larger than ± 3.3) and multivariate outliers (values with Mahalanobis distance greater than 26.125 - χ^2 [8] = 26.125, p <0.001) were removed. Following the analysis, only variables that significantly predicted variance in the dependent variable were included in the final model.

The final model for reading as shown in Table 14 includes four predictor variables, LLAMA B, LLAMA E, SRT and LLAMA D, which accounted for 30% of the variance in reading scores. Importantly, LLAMA B and E positively predict scores in reading, while the SRT task and LLAMA D are negative predictors.

Table 14. Hierarchical multiple regression model predicting reading proficiency

	Model 1		Model 2		Model 3		Model 4	
	B	β	B	β	B	β	B	β
Constant	76.186**		75.524**		79.176**		83.285**	
LLAMA B	0.175*	0.345	0.122	0.240	0.120	0.236	0.120*	0.237
LLAMA E			0.097*	0.271	0.089*	0.248	0.098*	0.273
SRT					-0.041*	-0.251	-0.042*	-0.256
LLAMA D							-0.117*	-0.234
R^2	0.119		0.182		0.244		0.298	
F	8.521**		6.875**		6.555**		6.326**	
ΔR^2	0.119		0.062		0.062		0.054	
ΔF	8.521 *		4.726*		5.023*		4.616*	

Note. * significant at 0.05 level; ** significant at 0.001 level

Table 15 shows the final model for the gap-fill test which includes two predictor variables explaining 19% of the variance in gap-fill scores: LLAMA F and forward digit span (FDS). In this model too there is both a positive predictor (LLAMA F) and a negative predictor (FDS).

Table 15. Hierarchical multiple regression model predicting L2 grammar knowledge

	Model 1		Model 2	
	B	β	B	β
Constant	68.851**		80.167**	
LLAMA F	0.165*	0.379	0.180	0.413
FDS			-0.154*	-0.225
R^2	0.144		0.193	
F	12.928**		9.096**	
ΔR^2	0.144		0.049	
ΔF	12.928 **		4.651*	

Note. * significant at 0.05 level; ** significant at 0.001 level

4.7. Discussion

The present study sought to contribute to our understanding of theory and measurement of the constructs of aptitude for explicit and implicit learning. To this end, we examined the relationship between LLAMA D and a probabilistic SRT task as hypothesized measures of implicit aptitude, and we scrutinized the relationship between all LLAMA subtests, the SRT and two measures of WM, that is, a construct that has been posited as another potential component of aptitude. Last but not least, we investigated the predictive power of aptitude for explicit and implicit learning and WM in relation to L2 proficiency, operationalized as grammar knowledge, reading, and listening comprehension. In the following, we discuss the findings in terms of their contribution to the conceptualization and operationalization of explicit and implicit aptitude in the field of L2 learning.

4.7.1. WM as a component of aptitude

Unlike most previous studies investigating WM in relation to aptitude, the present study included a comprehensive battery aimed at measuring not only explicit and implicit aptitude, but also both PWM and EWM. A factor analysis yielded separate factors for WM comprising PWM and EWM on the one hand, and explicit and implicit aptitude on the other hand, thus corroborating existing findings to the extent that comparisons can be made (Granena, 2013a; Hummel, 2009; Roehr & Gánem-Gutiérrez, 2009b; Yalçın & Spada, 2016). The cumulative evidence to date indicates that WM measures appear to tap a construct that is qualitatively different from aptitude, so research aimed at investigating variables interacting with and/or predicting L2 learning and use would ideally include measures of both aptitude and WM.

At the same time, we found a moderate correlation between EWM as operationalized via an operation span task and associative memory as measured by LLAMA B, a finding which suggests an involvement of executive function in the learning of new lexical items. Interestingly, we found no correlation between PWM operationalized via a forward digit span task and any of the LLAMA subtests. While similar results have been reported in other studies including measures of both EWM and PWM (Hummel, 2009), this result may seem counter-intuitive at first glance because PWM has been shown to be implicated in vocabulary acquisition (Juffs & Harrington, 2011; Li, 2016). However, if we take into consideration the factor of L2 proficiency, the finding is perhaps less surprising. Existing research suggests that the importance of PWM declines as proficiency increases (Hummel, 2009; Linck et al., 2013; Serafini & Sanz, 2016), and our advanced L2 learners may have crossed the threshold at which individual differences in PWM play a role. Indeed, the shared variance between the correlated measures of EWM and PWM in our study appears to confirm that it was the executive function component of WM that played a role in successful LLAMA B performance in the present study.

4.7.2. Implicit aptitude as a multi-componential construct

A factor analysis that included the LLAMA subtests, the SRT task and the two WM measures yielded three factors. LLAMA D and the SRT task as hypothesized measures of implicit aptitude loaded on the same factor and separately from measures of explicit aptitude and WM. This result substantiates the argument that implicit and explicit aptitude are separate constructs, each comprising distinct underlying abilities (Li & DeKeyser, 2021), and that LLAMA D and the SRT task tap abilities that are part of the same construct of implicit aptitude (Granena, 2013a, 2016). However, this finding needs to be considered in conjunction with another, seemingly contradictory result, that is, the absence of a correlation between LLAMA D and the SRT task. A possible explanation that immediately suggests itself is the difference in modality between the two tests. The SRT task is visual in nature, whereas LLAMA D is an auditory task. Research in cognitive psychology has shown that sensory modality can constrain higher-level cognition, including learning and memory (Conway et al., 2009). Moreover, the respective accuracy of auditory versus visual pattern perception may not be comparable (Collier & Logan, 2000). Furthermore, the SRT task and LLAMA D differ in terms of stimulus domain, with the former relying on non-verbal and the latter on verbal stimuli. Findings from neurocognitive research suggest numerous neurophysiological differences in the processing of verbal as opposed to non-verbal stimuli (Gevins et al., 1995). The SRT task is an RT measure that gauges the process of on-task learning (i.e., it is a processing-based measure, Christiansen, 2019). By contrast, LLAMA D is an accuracy measure which assesses learning offline (Christiansen, 2019). Thus, the SRT task measures the process of learning, whereas LLAMA D measures the product of learning.

Having said this, the fact that the SRT task and LLAMA D differ in terms of sensory modality, draw on different stimulus domains, assess the process versus the product of learning and are not statistically associated does not necessarily mean that they cannot be part of the same construct. Indeed, a lack of correlation between assumed measures of implicit aptitude has been reported in several recent studies (Buffington et al., 2021; Godfroid & Kim, 2021; Li & Qian, 2021). If the primary abilities involved in implicit aptitude are relatively disparate in nature, a lack of association between measures tapping these primary abilities would be less surprising. This suggestion is supported by DeKeyser and Li (2021), who have argued that implicit learning may occur via diverse pathways, and therefore abilities tested by implicit aptitude measures may not necessarily overlap or even intersect. In other words, implicit aptitude may be a multi-componential construct (Godfroid & Kim, 2021; Granena, 2020; Li & DeKeyser, 2021).

This line of argument is further supported by the fact that even though LLAMA D and the SRT task loaded on the same factor, the SRT task loaded positively and LLAMA D negatively on that factor. A recent study (Iizuka & DeKeyser, 2023) investigating the effect of different types of LLAMA D instructions reported a similar finding when participants were instructed to 'just listen' to the sound sequences in LLAMA D. In that condition, LLAMA D and SRT scores were negatively correlated (i.e., the abilities measured by these two tests were pulling in opposite directions). In the present study, participants were instructed to listen carefully to the sound sequences, and they were also told that they would be tested subsequently. Our instructions were thus different and arguably more explicit; the abilities measured by LLAMA D and the SRT task likewise pulled in opposite directions.

A possible explanation for this pattern of results is that participants may be approaching both LLAMA D and the SRT task in the same way, in line with their individual proclivities and regardless of the instructions they are given. In other words, they employ the same set of abilities on all versions of LLAMA D and on the SRT task, but due to the distinct nature of these two tests, such an approach has a facilitative effect in one case and a debilitative effect in the other. Specifically, Iizuka and DeKeyser (2023) suggest that focal attention may facilitate performance on LLAMA D but hinder performance on the SRT task. Success on the latter may depend on "the degree to which one is able to let go of the tendency to look for patterns and process input without focal attention" (Iizuka & DeKeyser, 2023, p. 19). Along similar lines, Kaufman et al. (2010) have suggested that, among other factors, openness and intuition are associated with success on the SRT task.

These considerations arguably shed new light on the construct of aptitude more generally because they imply that more is not necessarily better. If implicit aptitude in particular were not an ability in the classic sense (i.e., higher levels are invariably advantageous), but rather a propensity (see also Granena, 2016), where reliance on the right capacity at the right time and in the right context determines success, then this would no doubt change the outlook of L2 researchers, L2 teachers and L2 learners alike. At this point, such a line of argument is admittedly speculative. However, we believe it can usefully inform further research into the interrelations between different measures of (implicit) aptitude and their predictive power with regard to different components of L2 proficiency.

4.7.3. Predictors of L2 proficiency at advanced levels

Through a hierarchical multiple regression analysis, we identified predictors for two of our three proficiency measures. Overall, both explicit and implicit abilities predicted L2 morphosyntactic knowledge and L2 reading comprehension, in line with previously reported findings (Li, 2015). More specifically, LLAMA F positively predicted L2 morphosyntactic knowledge, while PWM as measured by a forward digit span task was a negative predictor, with a total of 19% of the variance accounted for. Moreover, LLAMA B and E positively predicted L2 reading, while the SRT task and LLAMA D were negative predictors, with a total of 30% of the variance explained.

Taking the latter finding first, we can see that two components of explicit aptitude, associative memory as measured by a vocabulary learning task and phonetic coding ability as measured by a sound-symbol association task, predicted performance on a reading test that assesses knowledge of language form and meaning, implied meaning and reading comprehension. This is entirely in line with expectations: Grapheme-phoneme mappings (or sound-symbol correspondence) and lexical knowledge are the very foundations of reading skill (in an alphabetic language). More strikingly, the two hypothesized measures of implicit aptitude used in the present study, the SRT task and LLAMA D, proved to be significant negative predictors. Put differently, domain-general implicit sequence learning ability (SRT task) and auditory pattern recognition ability (LLAMA D) were disadvantageous for reading performance, if relied upon solely (given that explicit aptitude components were already accounted for in the model).

With regard to morphosyntactic knowledge, language-analytic ability as measured by LLAMA F was a significant predictor. This is not only in line with previous empirical research (Li, 2015, 2016, 2022), but also theoretically coherent, since language-analytic ability can be expected to be important for the acquisition of grammar. As in the case of reading, the regression analysis for morphosyntactic knowledge also yielded a negative predictor, that is, PWM as measured by a forward digit span task. A similar argument as in the case of reading can be put forward: Exclusive reliance on phonological storage and rehearsal works against successful performance on a gap-fill task targeting selected linguistic structures (given that language-analytic ability was already accounted for in the model).

Finally, it is worth noting that none of the aptitude or WM measures included in the present study predicted L2 listening, our third measure of proficiency. As L2 listening was correlated with both L2 reading and L2 morphosyntactic knowledge, it is possible to conjecture that these latter two skills may have functioned as mediators. In other words, learners invested their explicit aptitude in acquiring morphosyntactic knowledge and reading skill, whereas listening skill was developed on the back of these. While this proposed explanation must remain speculative, it does sit well with the context in which the present study was conducted (i.e., an English-as-a-foreign-language setting characterized by form-focused classroom instruction that heavily relies on metalinguistic and literacy skills).

As in the case of the role of WM discussed above, the findings relating to predictors of L2 grammar, reading and listening highlight the importance of not only the learning context, but also learners' prior language learning experience in the sense of their proficiency level at the time of testing. Different constellations of cognitive (and other) variables can be expected to play different roles at beginner, intermediate and more advanced levels. Therefore, it is crucial to bear in mind that results from more advanced participants as reported here may not be generalizable to learners at lower levels of proficiency, and vice versa.

4.8. Conclusion

The present study measured explicit and implicit aptitude and WM in a group of L2 English learners of relatively advanced proficiency. Our empirical results corroborate a conceptual differentiation between explicit and implicit aptitude on the one hand and WM on the other hand, which suggests that the use of separate measures for these constructs is advisable.

In theoretical terms, the findings that (1) the hypothesized components of implicit aptitude pulled in different directions and (2) implicit aptitude components and PWM were negative predictors of L2 grammar and reading skill encourage us to consider the possibility that (implicit) aptitude may be a cognitive proclivity rather than an ability of immutable, context-independent value. This argument is in alignment with a comment put forward by

(Iizuka & DeKeyser, 2023, p. 17) in which they refer to aptitude considered in this way as being reminiscent of (cognitive or learning) style (see also Granena, 2016). It also chimes with earlier research taking a multi-dimensional and dynamic view of aptitude (Robinson, 2005, 2007, 2012), according to which the sensitivity of aptitude to environmental factors is such that it can be either activated or inhibited, based on the characteristics of various learning conditions. Over and above the role of learning context, our findings have highlighted the role of learners' proficiency level in the aptitude-outcome equation.

Despite yielding valuable insights, the present study was not without limitations. In particular, a limited number of exclusively cognitive variables was measured. Moreover, a larger sample size would have been desirable because this would have allowed for the empirical corroboration or otherwise of the currently entirely speculative argument that grammar and reading skill mediated the subsequent acquisition of listening skill.

In line with the findings reported here and in other recent studies on explicit and implicit aptitude, future research seeking to track the changing roles and relative weights of cognitive predictors as L2 proficiency develops would be of great value. In addition, the conceptualization of aptitude as a proclivity rather than as a fixed, context-independent ability deserves consideration in both the empirical and the theoretical domain, hopefully leading to well-informed research designs that capture multiple variables characterizing learners and the learning context. Last but not least, research aimed at identifying predictors would ideally draw on an experimental design where not only the product of learning, but also the process of learning is subject to experimental control and thus more ready interpretability. Work within an aptitude-treatment interaction paradigm (e.g., DeKeyser, 2021) would satisfy these criteria.

5. Chapter 5 (Article 2) – Finding the tipping point: Proficiency as a mediator of aptitude in L2 learning

5.1. Abstract

Research suggests that the role of language aptitude and working memory in L2 learning is dependent on learners' overall level of L2 proficiency – a construct typically operationalized in terms of categories. In the present study, we investigated the mediating role of proficiency, operationalized as a continuous variable, with a view to identifying the precise tipping point at which the role of specific cognitive abilities changes from facilitative to neutral, or vice versa.

L1 Croatian participants (N=86) completed the LLAMA aptitude test suite, a serial reaction time task, and measures of phonological and executive working memory, as well as outcome measures focusing on selected English morphosyntactic structures: a self-paced reading task, oral elicited imitation and gap-fill tests. L2 proficiency was operationalized as speaking skill in terms of complexity, accuracy and fluency.

Multilevel modelling confirmed the predictive power of explicit and implicit aptitude and working memory for L2 knowledge of the target structures. Proficiency was identified as a mediator: beneficial effects of explicit aptitude decreased significantly as proficiency increased, while the facilitative effects of implicit aptitude came to the fore, though non-significantly. Crucially, we identified the exact point on the proficiency continuum at which explicit aptitude loses its significant positive impact.

5.2. Introduction

The field of additional or second language (L2) learning is seeing renewed interest in the role of cognitive ability factors such as language learning aptitude and working memory. While numerous studies have affirmed the predictive power of aptitude in general terms (Li, 2015, 2016; Pavlekovic & Roehr-Brackin, 2024) [Chapter 4], evidence concerning the specific learning stage(s) at which different aptitude components predict achievement is not unequivocal. Some studies show that aptitude is a significant predictor in the early stages of L2 learning (Artieda & Muñoz, 2016), whereas others suggest that it is important at more advanced levels (Linck et al., 2013).

Recent theoretical developments have led to a distinction between aptitude for explicit and aptitude for implicit learning, with a number of studies supporting this conceptual contrast (Granena, 2012, 2013a; Pavlekovic & Roehr-Brackin, 2024; Roehr-Brackin et al., 2023; Suzuki & DeKeyser, 2017) [Chapter 4]. However, it remains unclear to what extent explicit and/or implicit aptitude facilitate L2 achievement differently at different proficiency levels. This is because the majority of existing studies on aptitude only included measures of explicit aptitude (Li, 2015) or aptitude was operationalized as a unidimensional construct (overall aptitude scores in Li, 2015), despite evidence pointing towards a multi-componential nature of both explicit and implicit aptitude. In addition, researchers have posited a predictive role for working memory, but again, evidence for its influence at different learning stages is contradictory (Kormos & Sáfár, 2008; Linck et al., 2013; Serafini & Sanz, 2016).

Beyond conceptual and methodological differences between empirical studies, there is a tendency to rely on statistical techniques which treat the participant sample as homogeneous and proficiency as a categorical variable, even though it is measured on a continuum. As a consequence, no study to date has been able to pinpoint the threshold at which the role of proficiency in the aptitude-outcome equation shifts from facilitative to neutral, or vice versa.

The present study addresses these issues by investigating the effects of explicit and implicit aptitude and working memory on L2 achievement in terms of explicit, automatized explicit and implicit knowledge of selected morphosyntactic structures. Not only do we examine the mediating role of proficiency in this relationship, but we also treat proficiency as a continuous variable. Employing a multilevel modelling approach, we aimed to identify the tipping point at which different aptitude components lose or gain their facilitative influence on L2 outcomes.

5.3. Background

5.3.1. Language learning aptitude

Language learning aptitude encompasses cognitive and perceptual abilities that facilitate the rapid and effortless learning of new languages (Carroll, 1981). The classic model of aptitude (Carroll, 1981) comprises phonetic coding ability, associative memory, and language-analytic ability (Skehan, 1998). This model is the result of a componential approach and was developed a posteriori on the basis of empirical results. In current research, these abilities are conceptualised as components of aptitude for explicit learning (hereafter: explicit aptitude). Explicit aptitude positively predicts learning outcomes (Li, 2015, 2016), that is, higher levels of aptitude are advantageous for L2 learning.

In addition to explicit aptitude, current research has posited the complementary notion of aptitude for implicit learning (hereafter: implicit aptitude) (Li & DeKeyser, 2021). It has been suggested that implicit aptitude comprises implicit learning ability and implicit learning memory (Granena, 2020). The cognitive abilities involved are sensitivity to frequency and conditional probability, priming or the tendency to be influenced by recent events, and selective attention (Li & DeKeyser, 2021). Similar to explicit aptitude, implicit aptitude is thus componential in nature. Factor analyses have yielded varied results, however. Specifically, implicit aptitude measures such as sequence learning ability and auditory pattern recognition have been found to both load positively onto the same factor (Granena, 2012, 2013a; Roehr-Brackin et al., 2023), or one measure has been found to load positively and the other negatively onto the same factor (Iizuka & DeKeyser, 2023; Pavlekovic & Roehr-Brackin, 2024) [Chapter 4].

Theoretical models of aptitude informed by an information-processing paradigm have focused on explicit aptitude, although the role of implicit aptitude is acknowledged in more recent work. The staged model of aptitude (Skehan, 2002, 2012) contends that different aptitude components hold varying degrees of importance at different learning stages. For instance, phonetic coding or "handling sound" (Skehan, 2019, p. 61) is particularly relevant at the initial stage. Language-analytic ability becomes crucial in the subsequent "handling pattern" stage, where learners identify and generalize linguistic patterns. Automatization, on the other hand, is most relevant in the final stage, "automatising-proceduralizing" (Skehan, 2019, p. 61). Both Skehan (2016) and Li (2022) propose that the automatization process is heavily reliant on implicit learning abilities. Taken together, this would imply that cognitive abilities facilitating explicit learning are more relevant at the initial stages of L2 acquisition, whereas cognitive abilities underlying implicit learning become more important at advanced stages. Working memory is deemed important from the initial stages onwards, with retrieval memory taking over in later stages. Along similar lines, the model of aptitude complexes proposed by (Robinson, 2005, 2012) posits that different clusters of cognitive abilities are engaged in different learning and instructional contexts.

5.3.2. Working memory

The aptitude models reviewed in the preceding section acknowledge the role of working memory (WM) as an individual difference factor predicting L2 outcomes (Robinson, 2005, 2007, 2012; Skehan, 2002, 2012, 2016). WM refers to the ability to concurrently store and process information during cognitive tasks (Baddeley, 1986; Cowan, 2005). Baddeley's multi-componential model of WM has been influential in L2 learning research (Baddeley, 2015, 2017; Wen, 2019), which has particularly focused on the components of phonological working memory (PWM) and executive working memory (EWM). PWM is responsible for the short-term storage of phonological information and articulatory rehearsal, while processes such as inhibition, updating, and switching are part of EWM (Wen, 2019). The importance of PWM and EWM in the processing, learning, and use of an L2 is widely accepted (Juffs & Harrington, 2011; Linck et al., 2014). Research tends to support a 'the-more-the-better' hypothesis, suggesting that learners with higher WM will outperform those with lower WM (Miyake & Friedman, 1998). However, conflicting findings exist, with some studies showing no significant role of WM (Juffs, 2004, 2005). Discrepancies in findings can arise from different WM components being investigated, differences in measurement techniques, and factors such as characteristics of the outcome measures used (Serafini & Sanz, 2016). Therefore, the precise role of (components of) WM at different learning stages remains unclear.

5.3.3. Empirical evidence for the role of aptitude and working memory at different levels of L2 proficiency

Studies investigating the effects of aptitude and WM have found that facilitative effects depend on learners' overall L2 proficiency, with recent research suggesting that explicit aptitude is more crucial at initial stages of L2 learning, whereas implicit aptitude plays a greater role at advanced stages (Li & DeKeyser, 2021).

Linck et al. (2013) used the Hi-LAB battery to investigate the role of cognitive and perceptual abilities at advanced proficiency levels. They categorized participants into mixed-attainment and high-attainment groups based on their scores on a Defense Language Proficiency Test or by the demands of their occupations. The findings revealed that implicit aptitude as measured by a serial reaction time task, associative memory as assessed by a paired associates test, and PWM as measured by a letter span task were all positively correlated with high language attainment. Surprisingly, the researchers also noted a negative correlation between switching ability, a component of EWM measured by a task switching test, and high attainment. They attributed this to more pronounced switching between L2 and L1 and consequently greater reliance on L1, which in turn would hinder high attainment in the L2 (Linck et al., 2013). Together, this suggests that both explicit and implicit aptitude as well as WM play predictive roles in language success at advanced proficiency levels, although some components – such as an EWM measure in this case – may emerge as negative predictors.

Artieda and Muñoz (2016) included beginner- and intermediate-level learners in their study. They examined the relevance of aptitude as assessed by the LLAMA battery (Meara, 2005; Meara & Rogers, 2019) at these two proficiency levels. Results revealed that when treated as a unidimensional construct, aptitude exhibited a consistent medium effect size at both beginner and intermediate levels of proficiency. When sub-test scores were analysed separately, auditory pattern recognition ability had a significant impact at beginner level only. Conversely, language-analytic ability was relevant at both beginner and intermediate level, albeit more so at intermediate level. Associative memory did not play a role at either proficiency level.

A meta-analysis of 66 aptitude studies (Li, 2016) demonstrated that different aptitude components predict various L2 skills, including listening, reading, speaking, grammar, and overall L2 knowledge. Phonetic coding ability emerged as a strong predictor of overall proficiency, particularly in beginner learners, pointing towards its greater relevance at initial stages of L2 learning. Language-analytic ability strongly predicted grammar learning and reading comprehension, while associative memory showed a weak relationship with L2 proficiency. Moreover, the meta-analysis revealed that explicit aptitude was more predictive of L2 achievement in high-school learners than university students. Considering that high-school participants typically exhibited lower proficiency compared with university students, this implies that explicit aptitude is more relevant for beginners, consistent with Carroll's comment that classic aptitude pertains to learning a language "from scratch" (Carroll, 1990, p. 24).

Working with learners at an advanced proficiency level, Suzuki and DeKeyser (2017) examined the predictive relationship between auditory pattern recognition, language-analytic ability, and PWM with specific types of L2 knowledge, that is, implicit and automatized explicit knowledge, each operationalized via three different tests targeting structures of Japanese. The findings provided evidence for a significant relationship between language-analytic ability and automatized explicit knowledge.

Some researchers specifically examined the role of WM at different learning stages. In a study investigating the impact of WM at three proficiency levels (Serafini & Sanz, 2016), beginner, intermediate, and advanced learners were assessed in terms of their L2 morphosyntactic development. Results revealed that PWM as measured by a digit span task was a significant predictor on an oral elicited imitation task in the beginner and intermediate groups. However, no such relationship was observed in advanced learners. Similarly, on an untimed grammaticality judgement task, PWM showed a significant relationship for intermediate learners only. EWM as assessed by an operation span task did not emerge as a significant predictor in a regression analysis, but it showed a weak to moderate positive correlation with the elicited imitation task in a simple bivariate analysis for beginner and intermediate learners.

Kormos and Sáfár (2008) explored the relationship between WM and various L2 skills, including reading, listening, speaking, and overall L2 proficiency, among beginner and intermediate learners. Results showed that PWM as measured by a non-word span test was significantly related to overall proficiency, English language use, and a composition task in the intermediate group. Conversely, EWM as measured by a backward digit span task predicted overall proficiency scores and achievement in L2 reading, listening, and speaking in the beginner group.

Taken together, the empirical studies reviewed in the preceding paragraphs point towards a mediating role for L2 proficiency in the effects of aptitude and working memory on L2 outcomes. At the same time, findings are not necessarily convergent. Contrasting results can be due to the use of different measures of aptitude and WM. However, they may also be attributable to the conceptualisation of proficiency in terms of two or three broad categories. Furthermore, L2 achievement has been operationalized in various ways. As current research posits explicit and implicit aptitude, it would arguably be theoretically most coherent to conceptualize L2 outcomes with reference to the complementary notions of explicit and implicit knowledge.

5.3.4. Explicit and implicit knowledge and learning

Implicit learning pertains to the acquisition of implicit or unconscious knowledge without the intention to learn or awareness of the knowledge acquired. By contrast, explicit learning involves the acquisition of conscious or declarative knowledge under intentional learning conditions (Rebuschat, 2013). Implicit knowledge is characterized as tacit, intuitive, and beyond conscious introspection; it typically involves rapid access (Hulstijn, 2015; Rebuschat, 2013). Explicit knowledge is conscious, available for introspection, and potentially verbalizable (N. Ellis, 2015; Rebuschat, 2013).

Tasks that limit conscious reflection and require language processing in real time are hypothesized to access implicit knowledge (Suzuki & DeKeyser, 2015; Vafaee et al., 2017). Self-paced reading and word monitoring tasks are examples of such measures, where sensitivity to grammatical errors as measured by reaction times is considered an indicator of implicit knowledge of the targeted morphosyntax.

Some studies (R. Ellis, 2005; Godfroid & Kim, 2021) have supported elicited imitation as a measure of implicit knowledge. By contrast, Suzuki and DeKeyser (2015) argued that elicited imitation tasks measure neither implicit nor explicit knowledge, but a type of knowledge that shares characteristics of both. They propose the notion of automatized explicit knowledge – a type of knowledge that is similar to implicit knowledge in that it involves fast access to linguistic representations, and also to explicit knowledge because learners are asked to repeat sentences correctly (Suzuki & DeKeyser, 2015). This view is supported by recent neurolinguistic evidence, especially with regard to the production (imitation) phase of the task (Suzuki et al., 2023).

Explicit knowledge is typically gauged via accuracy-focused tasks which allow for conscious reflection. Examples are gap-fill tests in which participants select a correct answer from a number of given options and tests of metalinguistic knowledge (R. Ellis, 2005) in which participants are required to supply both a correct answer and the underlying metalinguistic rule.

5.4. The current study

In summary, existing research has shown that explicit aptitude is important for L2 learning at lower levels of proficiency, whereas its significance at later stages is less clear. Conversely, it has been proposed that implicit aptitude plays a facilitative role at advanced levels, though empirical evidence is still in short supply. The role of WM at different proficiency levels is unclear. PWM may be particularly important at initial stages, while results pertaining to EWM are contradictory, with some studies indicating its significance for beginners and others for advanced learners.

To date, studies investigating aptitude and WM at different learning stages have treated proficiency as a categorical variable, that is, participants are classified as 'beginners', 'intermediate' or 'advanced', with a range of proficiency scores included in each category

(Artieda & Muñoz, 2016; Serafini & Sanz, 2016). While this is certainly a possible approach, it does not allow us to determine at what point on the proficiency continuum specific cognitive abilities become, or cease to be, relevant – an issue which is of critical importance both for a theoretically driven research agenda and for practical purposes. Accordingly, the present study addressed the following research questions:

RQ4: To what extent are the effects of explicit and implicit aptitude and working memory on learners' knowledge of selected L2 morphosyntactic structures mediated by L2 proficiency?

RQ5: At which point on the L2 proficiency continuum do the facilitative effects of explicit and implicit aptitude and working memory become or cease to be significant?

5.5. Methodology

The current study had a correlational design with online administration of all measures. Explicit and implicit aptitude were assessed by means of the LLAMA test suite and a probabilistic serial reaction time task. Working memory was measured via forward digit and operation span tasks. The L2 outcome measures focused on three morphosyntactic structures assessed by means of a self-paced reading task to capture implicit knowledge, an elicited imitation test to capture automatized explicit knowledge, and a gap-fill test to capture explicit knowledge. L2 proficiency was evaluated by means of a monologic oral production task which was analysed in terms of complexity, accuracy, and fluency. In the subsequent sections, the study participants, instruments, and procedure are described in full.

5.5.1. Participants

The study involved 86 L1 Croatian learners of English who had been exposed to L2 instruction for between 6 and 13 years (*M* = 10, *SD* = 1.72) as part of their compulsory secondary-school education. The sample consisted of 62 female and 22 male participants as well as 2 participants who opted not to disclose their gender. The learners' age ranged from 15 to 18 years (*M* = 16.14, *SD* = 1.29). All participants had been studying English as a compulsory subject, receiving three hours of instruction per week since the beginning of secondary school. Classes focus on communicative language skills but also include form-focused instruction to prepare students for the nationwide exam at the end of their final year. A total of 68 participants (81.9%) had never visited an English-speaking country, while 2 participants (2.4%) had stayed abroad for a year. The remaining 13 participants (15.7%) reported short stays averaging 25 days.

5.5.2. Instruments

5.5.2.1. L2 outcome measures: implicit, automatized explicit, and explicit L2 knowledge

The L2 outcome measure comprised three tasks aimed at respectively capturing implicit, automatized explicit, and explicit knowledge of three selected morphosyntactic structures.

Implicit knowledge was assessed by means of a self-paced reading task (Jiang, 2007) using a moving-window presentation. Participants were required to read sentences word by word and as quickly as possible while focusing on their meaning to accurately answer comprehension questions. Each sentence began with the first word displayed on the left-hand side of the computer screen. Upon pressing a designated key, the word disappeared, and the next word appeared to the right. Half of the target sentences were grammatically correct, and half contained a grammatical violation.

Three regions of interest (ROIs) were embedded in the task: (1) the critical word where the grammatical violation could be detected, (2) the word following the critical word, and (3) the word following the word at ROI 2. A delay in reaction time (RT) was expected for ungrammatical sentences at ROI 1, while ROIs 2 and 3 captured any spillover effects.

The stimuli consisted of 72 target sentences encompassing the three target structures and 24 fillers; the latter were all grammatically correct. Each sentence was followed by a yes/no comprehension question, with an equal ratio of positive and negative responses. Sentences testing the same structure did not appear consecutively more than twice. A grammatical sensitivity index was calculated by subtracting RTs for grammatically correct sentences from RTs for ungrammatical sentences at ROIs 1, 2, and 3 combined (Suzuki, 2017).

Automatized explicit knowledge was assessed via an oral elicited imitation test (R. Ellis, 2005; Erlam, 2006; Godfroid & Kim, 2021; Suzuki & DeKeyser, 2015). Participants listened to a series of sentences, each followed by five random numbers. They were instructed to (1) listen to the sentence, (2) read aloud each number as it appeared, and (3) repeat the sentence in correct English. As participants had to attend to the numerical stimuli, rehearsal and rote repetition of the target sentences was prevented (Mackey & Gass, 2022). Participants had a 9-second time limit for sentence repetition, which had been determined through piloting.

There were 72 sentences covering the three target structures, with each set comprising an equal number of grammatically correct and grammatically incorrect sentences. Sentences testing the same structure did not appear more than twice in succession. The task did not include fillers to reduce participant fatigue. Piloting with 22 participants followed by post-task interviews had shown that test takers had not identified the target structures. Sentence lengths ranged from 8 to 24 syllables, with a mean length of 15.68 syllables. Mean sentence duration was 5.53 seconds, with the shortest at 3.9 seconds. This duration exceeds the typical time span of 1.5-2.0 seconds for information to decay in phonological short-term memory (Baddeley et al., 1975). The test was scored for accurate production of the targeted structures. Avoidance was scored as zero; errors that did not pertain to the targets were ignored. The maximum score was 72.

Explicit knowledge was measured through a gap-fill test comprising 75 sentences targeting the three morphosyntactic structures and employing a three-way multiple-choice answer format from which participants were required to choose one correct option in each case. The maximum score was 75.

Items from the L2 outcome measures can be found in Appendix 1.

5.5.2.2. Target structures

The above tasks targeted three morphosyntactic structures: past simple tense, passive voice, and articles. These structures feature prominently in the participants' grammar syllabus and are the source of frequent errors, as identified by the learners' English teacher (D. Linić Učur, personal communication, 06/07/2020). Therefore, the structures were a suitable focus for measures aimed at achieving maximum discrimination between participants.

5.5.2.3. Measures of explicit and implicit aptitude

Language learning aptitude was assessed by means of the LLAMA suite (Meara, 2005; Meara & Rogers, 2019) and a probabilistic serial reaction time (SRT) task (Kaufman et al., 2010). The LLAMA comprises four subtests, as follows.

LLAMA B (version 2) evaluates associative memory by requiring participants to learn 20 new vocabulary items for novel picture stimuli within a 2-minute learning phase. Subsequently, in an untimed test phase, participants are tasked with matching presented words to their corresponding pictures from a set of 20 options. The maximum score is 20, with 1 point awarded for each correct answer and no penalty for guessing.

LLAMA D (modified version 2) gauges auditory pattern recognition ability. During an exposure phase, participants hear 10 words in an unknown language. In the subsequent test phase, participants hear words from the same language, including both previously encountered and new items, and indicate familiarity through a yes/no response. The maximum score is 40 (20 new and 2x10 previously encountered items). Incorrect responses were penalised to compensate for guessing.

LLAMA E (modified version 3) measures sound-symbol correspondence. Participants are presented with 24 phonetic symbols, each representing a distinct syllable. Upon clicking on a symbol, the corresponding syllable is played. Participants can click on any symbol any number of times during a 2-minute learning phase. In the subsequent untimed test phase, participants hear two-syllable combinations and must select the correct answer from a set of 20 combinations of previously encountered symbols. A partial-credit scoring system was used, with a maximum score of 40. Up to two points were awarded for each correct answer - one point each for a correctly identified syllable and its correct position.

LLAMA F (modified version 3) is a grammatical inferencing task in which participants must work out the rules of an unknown language. During the 4-minute learning phase, participants interact with buttons revealing simple pictures accompanied by written descriptions. With 20 items in total, participants can click on any button as many times as they wish. In the subsequent untimed test phase which also comprises 20 items, participants are presented with similar pictures and must select combinations of words that accurately describe the given pictures. A partial-credit scoring system was used, with each correct word yielding up to two points: one point for identifying the correct word and one point for placing it in the correct position. The maximum score is 132.

The probabilistic SRT task requires participants to react to visual stimuli presented as black squares appearing in one of four possible locations on the computer screen. Stimulus sequences are generated according to a probabilistic rule, with 85% of the stimuli following a training sequence and the remaining 15% following a control sequence. The task comprises 8 blocks, each containing 120 trials, resulting in a total of 960 trials. Following the protocol outlined by (Kaufman et al., 2010), trials were initially randomized within each block and subsequently presented in the same fixed order for each participant. Task performance was assessed by calculating the difference in mean RT between training and control trials.

5.5.2.4. Measures of working memory

Phonological working memory (PWM) was measured by a forward digit span (FDS) task based on the format developed by Linck et al. (2013), which adapted a component of the operation span task by Unsworth et al. (2005). Participants listened to number sequences in their L1, ranging from three to nine digits in length, and were required to repeat back the numbers in order. Each set consisted of four sequences of a particular length, resulting in a total of seven sets and 28 sequences. The maximum score was 168.

Executive working memory (EWM) was assessed using an automated operation span (OSPAN) task developed by Unsworth et al. (2005). In this task, participants first solved simple mathematical problems and indicated whether the displayed solution was correct or incorrect. They were then presented with a letter to memorize. Following a sequence of mathematical problems and letters, participants were required to recall the letters in order. The task comprised 18 sets of sequences ranging from three to eight letters resulting in a total of 99 letters. The maximum score was 99.

5.5.2.5. Measure of L2 proficiency

Speaking proficiency was gauged via a monologic oral production task. Participants were given 3 minutes to talk without interruption about their typical day at school. Their speech was audio-recorded, transcribed, and analysed using the CLAN software (MacWhinney, 2000). The analysis yielded measures of complexity, accuracy, and fluency (CAF). Giraud's index was used to calculate lexical complexity (Guiraud, 1954); the number of clauses per c-unit was used to calculate morphosyntactic complexity (Foster et al., 2000). Lexical and morphosyntactic accuracy were operationalized as lexical and morphosyntactic errors per 100 words, respectively (adapted from Michel et al., 2007). Fluency was calculated through pruned speech rate, that is, the number of spoken words per minute (adapted from Freed, 2000).

5.5.3. Procedure

The oral production task was conducted via Zoom, while all other measures were programmed into PsychoPy and administered via the Pavlovia platform (Peirce et al., 2019). Test instructions were provided in L1. Participants were instructed to use headphones in a quiet setting while working at their own computers. The first author monitored the participants through Zoom to ensure protocol adherence and answer any clarification questions. Data were collected in three separate sessions, as shown in Figure 18.

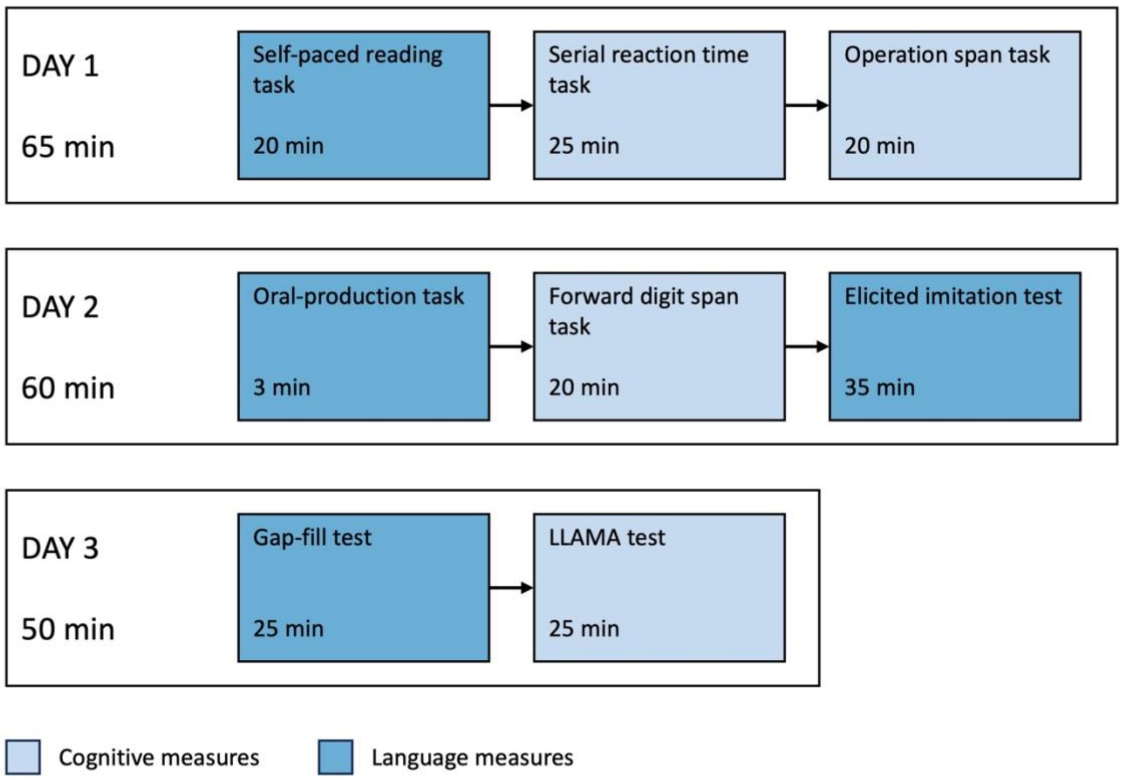

Figure 18. Procedure (Chapter 5)

5.5.4. Data analysis

To begin with, we computed reliability indices and scrutinized the data distributions. We calculated descriptive statistics for all variables. Bivariate correlations and exploratory factor analysis were used to explore the interrelationships and factoring of variables, respectively.

To address the research questions, we employed a multilevel modelling approach. Multilevel modelling allows for the inclusion of random effects, i.e. variables that are expected to generate random and thus unwanted variation. In the context of the present study, participants were included as random effects, enabling intercepts and slopes of predictor variables to vary across individuals (see Linck, 2016, for the benefits of using multilevel models for analysing individual differences; see Cunnings & Finlayson, 2015, for a detailed overview of multilevel modelling).

We examined main effects of aptitude and working memory, thus establishing their predictive power. In order to establish the mediating role of proficiency, we examined interactions between proficiency on the one hand and the relationship between cognitive ability (aptitude, WM) and L2 outcome measures on the other hand. Interaction terms and their significance were calculated and the slopes of the predictor-outcome relationship at different levels of proficiency were scrutinized.

Finally, we employed simple slopes analysis (Aiken et al., 1991; Rogosa, 1981) and Johnson-Neyman intervals (Johnson & Neyman, 1936) to determine the tipping point at which any effects of aptitude and WM changed from significant to non-significant or vice versa. Simple slopes analysis involves calculating the significance of a relationship between two variables at various (arbitrary) levels of the mediating variable. Johnson-Neyman intervals allow us to calculate the range of significance for a relationship and to visualize the results by showing the values and significance of the slope of a relationship across a continuous mediating variable, i.e. proficiency.

The significance level was set at .05. Reliability and normality checks along with factor analyses were performed in SPSS, v.29.0 (IBM Corp., 2023). Correlational analyses, multilevel modelling, and interaction analyses were carried out in R, version 2021.09.2 (2021), and version 4.3.2 (2024) . A full list of R packages used in Chapter 5 (Article 2) is provided in Appendix 3.

5.6. Results

This section presents the descriptive results for the variables under study followed by factor analyses. We then provide answers to our research questions on the basis of a multilevel modelling approach.

5.6.1. Preliminary analyses

Descriptive statistics were calculated for the L2 outcome measures, that is, the self-paced reading (SPR) task as a measure of implicit knowledge, the elicited imitation (EI) test as a measure of automatized explicit knowledge, and the gap-fill (GAP) test as a measure of explicit knowledge. In the SPR task, participants' mean RT for ungrammatical sentences (449ms) was longer than for grammatical sentences (439ms), as expected. A series of t-tests was conducted to investigate differences between grammatical and ungrammatical sentences. The GSI of 10ms for all three structures combined was significant with a small effect size, $t(82) = -3.596$, $p = .001$, $d = 0.28$ (Cohen, 1988). This suggests that participants were sensitive to grammatical violations, which, in turn, is indicative of implicit knowledge. It is also worth noting that participants scored relatively highly on both EI and GAP. The boxplots in Figure 19 provide an overview; full descriptive statistics can be found in Table A in Appendix 4.

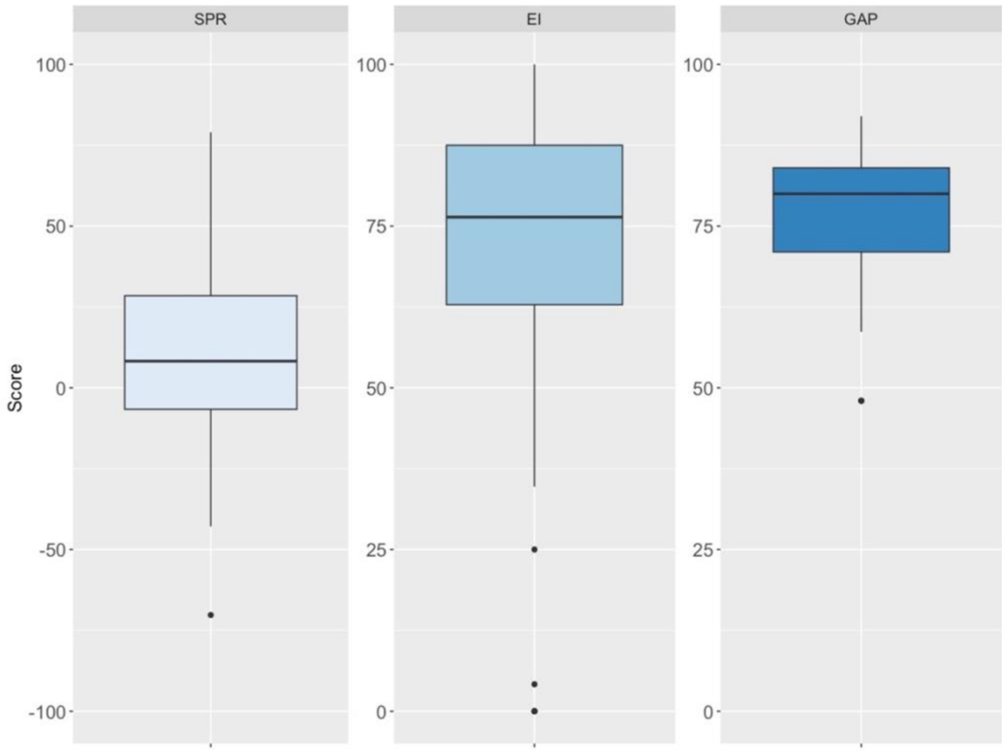

Figure 19. Boxplots of the outcome measures

Scores on the cognitive ability measures were adjusted to a scale out of 100. On the LLAMA tests, participants scored highest on LLAMA B, followed by F, E, and finally D. On the FDS and OSPAN, participants also scored highly. Boxplots are available in Figure 20; full descriptive statistics are available in Table A and Table E in Appendix 2.

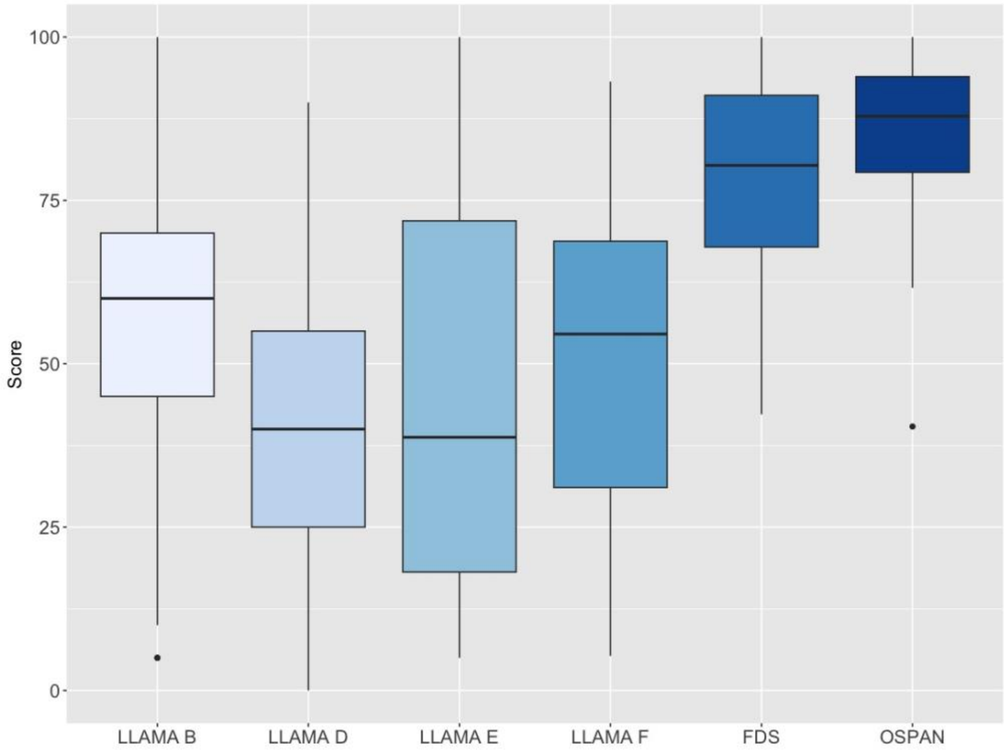

Figure 20. Boxplots of the cognitive variables

Participants' scores on the SRT task were computed and differences in mean RT across the 8 task blocks were examined to establish learning effects. Error responses (9% of the data) were excluded. Significant outliers, defined as values exceeding three standard deviations from the mean RT for each participant in each block (1% of the data), were also removed, resulting in a reduced sample size of 83. Each participant's SRT score was determined by subtracting the mean RT in the training condition (447ms, SD = 67) from that in the control condition (457ms, SD = 71). Mean RTs broken down by block are available in Table B in Appendix 2. A series of t-tests was conducted to assess learning effects (see Table C in Appendix 2). Differences between training and control conditions across blocks 4-8 were significant, $t(80) = 9.545$, p = .0001, with a medium effect size, $d = 0.75$ (Cohen, 1988), so data from these blocks were included in the final SRT score calculation. Split-half reliability with Spearman-Brown correction yielded a coefficient of 0.44 for blocks 4-8. This value is comparable to the reliability of similar tasks in previous studies (Granena, 2013b; Kaufman et al., 2010; Suzuki & DeKeyser, 2015, 2017).

The descriptive statistics for complexity, accuracy, and fluency as components of speaking proficiency are shown in Table 16.

Table 16. Descriptive statistics for proficiency variables

Measure	n	M	SD	Skew	S-W (*p*)
Lexical complexity	82	7.74	.86	.118	.870
Morphosyntactic complexity	82	1.59	.28	.681	.035
Lexical accuracy	82	1.02	.78	1.166	.001
Morphosyntactic accuracy	82	1.16	1.27	2.907	.001
Fluency	82	95.96	25.37	.263	.828

We conducted an exploratory factor analysis of the complexity, accuracy, and fluency measures, using principal component analysis with direct oblimin (oblique) rotation, given that the underlying factors were expected to be related (Tabachnick & Fidell, 2014, p. 651). Assumptions were met: KMO = 0.61; Bartlett's test of sphericity was significant ($p < .001$). The analysis yielded a single component with an eigenvalue above 1 ($\lambda = 2.64$) that accounted for 53% of the variance. We labelled this component 'speaking proficiency'. The factor loadings for each measure are shown in Figure 21; details can be found in Table B in Appendix 4.

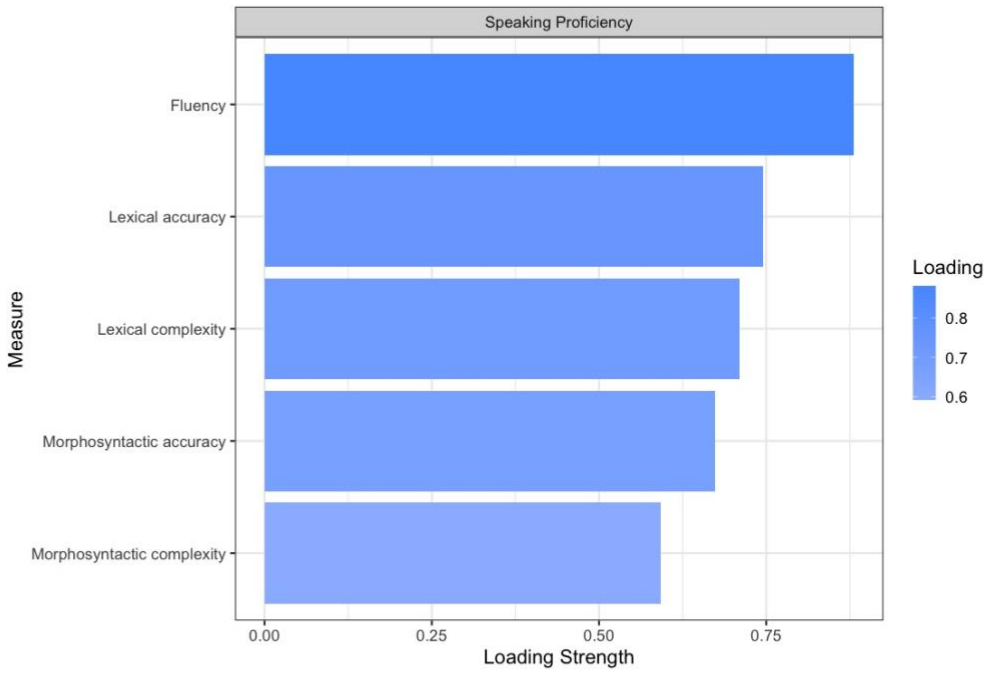

Figure 21. Factor loadings of speaking proficiency variables

Finally, we conducted an exploratory factor analysis to examine which cognitive ability measures would factor together. Again, we used principal component analysis with direct oblimin (oblique) rotation. Assumptions were met: KMO = 0.55; Bartlett's test of sphericity was significant ($p < .001$). The analysis yielded three components with eigenvalues above 1. LLAMA B, E, and F loaded on factor 1 ($\lambda = 1.97$), which accounted for 28% of the variance. OSPAN and FDS loaded on factor 2 ($\lambda = 1.31$), which explained 19% of the variance. Finally, the SRT task and LLAMA D loaded on factor 3 ($\lambda = 1.12$), albeit pulling in opposite directions, which explained 16% of the variance. Taken together, the three factors explained 63% of the variance. We labelled them (1) explicit aptitude, (2) working memory, and (3) implicit aptitude. The factor loadings are shown in Figure 22.

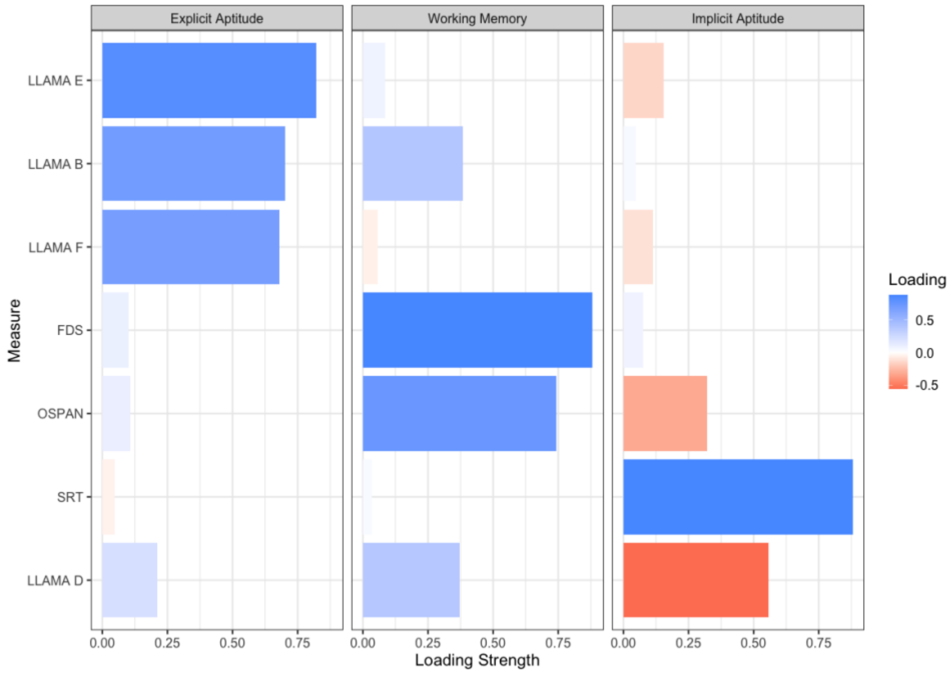

Figure 22. Factor loadings for a three-component solution (PCA) (for reference only; identical to Figure 17)

5.6.2. To what extent are the effects of explicit and implicit aptitude and working memory on learners' knowledge of selected L2 morphosyntactic structures mediated by L2 proficiency?

To address the first research question, interrelationships between the cognitive ability factors, the outcome measures and speaking proficiency were calculated. Figure 23 shows Spearman correlation coefficients (upper triangle), scatterplots for variable pairs (lower triangle) and the distribution of data for each variable (on the diagonal).

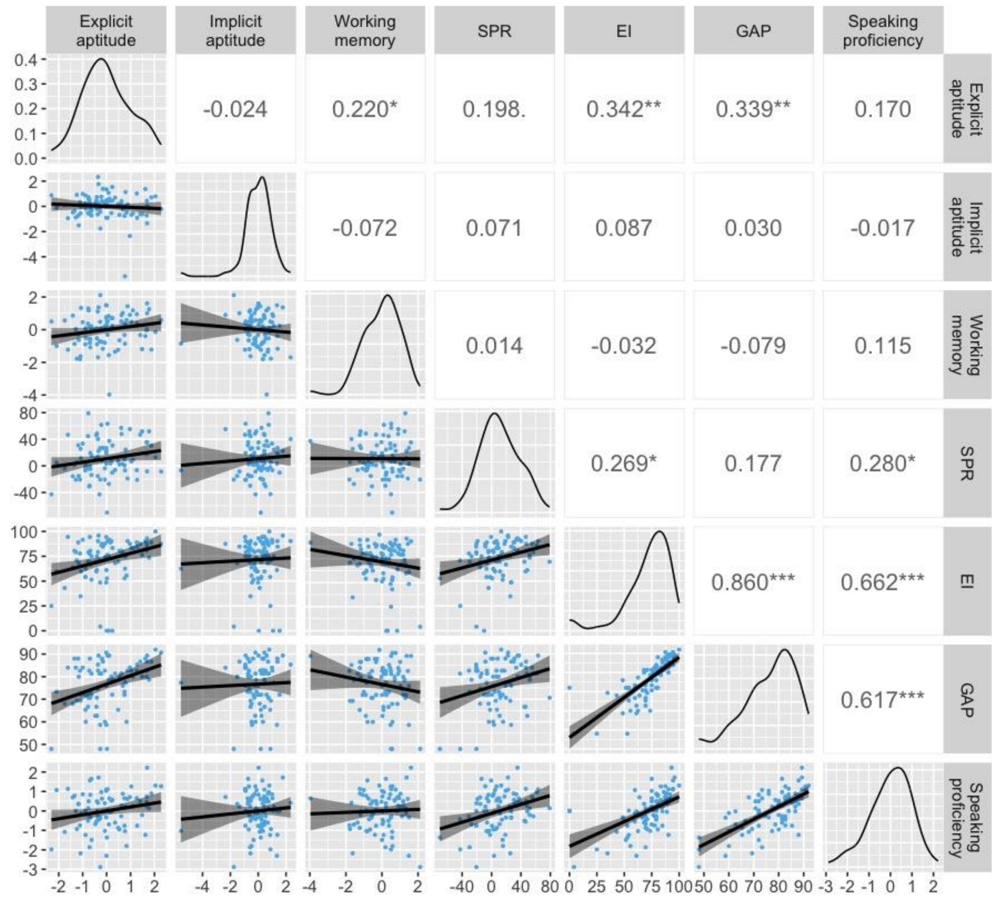

Figure 23. Correlations (Spearman's rho) between measures of aptitude and working memory, L2 knowledge, and L2 proficiency

As can be seen in Figure 23, explicit aptitude is positively and moderately related to EI (r_ρ = .342) and GAP (r_ρ = .339) scores. Conversely, WM and implicit aptitude are not related to any of the outcome measures. Among the cognitive ability factors, WM is positively related to explicit aptitude (r_ρ = .220).

To investigate the predictive power of explicit and implicit aptitude and WM, three multilevel models were built, that is, one model for each of the three outcome measures: SPR assessing implicit knowledge, EI assessing automatized explicit knowledge, and GAP assessing explicit knowledge. The linearity assumption was met for each predictor in each model. The models enabled us to identify main effects of explicit aptitude, implicit aptitude, and WM (predictors), as well as the interaction terms with proficiency (mediator). All predictors were mean-centred. Figure 24 provides a visual representation of the models; the source code and full details of the results are available in Appendix 5.

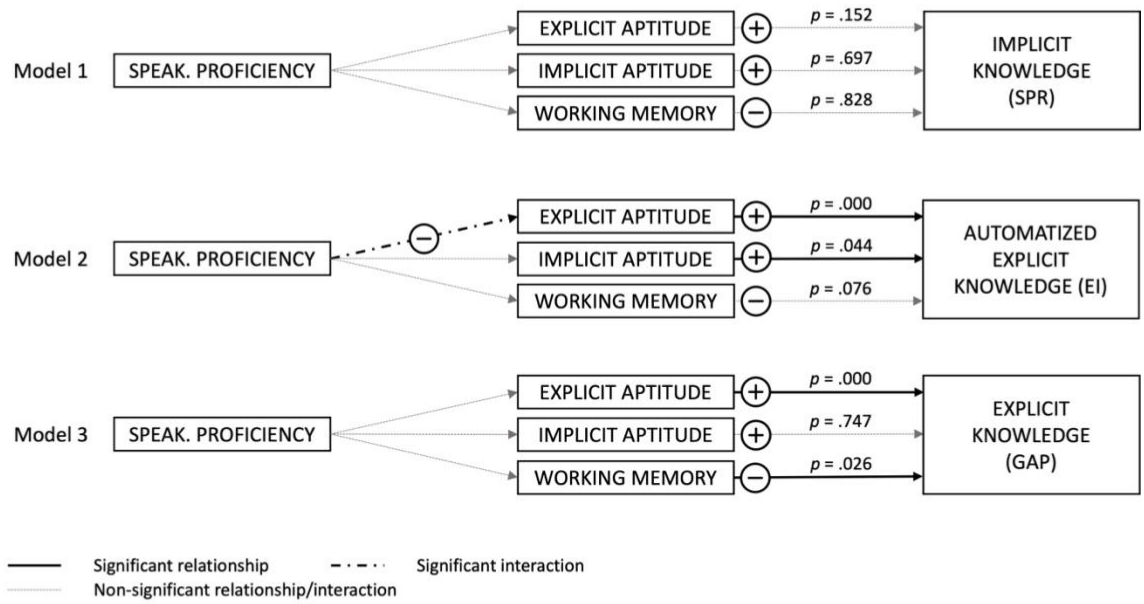

Figure 24. Multilevel models (Chapter 5)

In the case of automatized explicit knowledge (EI), a significant positive main effect of explicit aptitude (Model 2: estimate = 4.82, SE = 1.05, t = 4.597, p < .0001), and a significant positive main effect of implicit aptitude (Model 2: estimate = 2.12, SE = 1.04, t = 2.041, p = .044) were in evidence. In the case of explicit knowledge (GAP), a significant positive main effect of explicit aptitude was found (Model 3: estimate = 3.36, SE = 0.85, t = 3.977, p < .0001). Moreover, the analysis yielded a significant negative main effect of WM (Model 3: estimate = -2.19, SE = 0.86, t = -2.560, p = .026). There were no significant predictors for implicit knowledge (SPR, Model 1).

To investigate the mediating effects of proficiency, the three multilevel models were examined for significant interaction terms between proficiency on the one hand and the relationship between explicit and implicit aptitude and WM and the outcome measures on the other hand. The presence of such interactions would indicate that the relationship between predictor and outcome varies significantly across different proficiency levels. The interaction terms are illustrated in Figure 24, which shows a single significant negative interaction between explicit aptitude and proficiency in Model 2, predicting automatized explicit knowledge (EI) (estimate = -1.99, SE = 0.99, t = -1.997, p = .048). The negative interaction term indicates that the positive relationship between explicit aptitude and automatized explicit knowledge differs in strength across different levels of the mediator variable: at lower proficiency levels, explicit aptitude is a positive predictor of automatized explicit knowledge, but the relationship weakens as proficiency increases, making explicit aptitude a less robust predictor at higher proficiency levels. No other interactions were significant.

Our findings are visualised in more detail in Figure 8, which plots the relationships between explicit and implicit aptitude with the three outcome measures (y-axis) at five proficiency levels: mean proficiency -2SD, mean proficiency -1SD, mean proficiency, mean proficiency +1SD, and mean proficiency +2SD (x-axis). Following the lines from left to right in each graph, we can see that they do not run in parallel, but gradually diverge or converge instead; this indicates that the slope of the relationship is different at different proficiency levels. The middle plot on the left shows that the slope is much steeper at lower proficiency levels (e.g. M-2SD) compared to higher proficiency levels (e.g. M+2SD), that is, the relationship loses power as proficiency increases. This is the significant interaction reported above: proficiency negatively mediates the effect of explicit aptitude on automatized explicit knowledge.

Interestingly, a similar visual pattern can be observed in the top and bottom plots on the left-hand side, despite the absence of significant effects of explicit aptitude in Model 1 or significant interaction terms in Models 1 and 3. Overall, the pattern suggests that in the case of implicit and explicit knowledge too, the facilitative effect of explicit aptitude diminishes as proficiency increases.

The plots on the right-hand side of Figure 25 focus on implicit aptitude and show the opposite pattern. Despite a lack of significance, the trend is particularly clear in the middle plot which displays the relationship between implicit aptitude and automatized explicit knowledge. At lower proficiency levels, the slope is near zero, as indicated by an almost horizontal line; in other words, there is no relationship between implicit aptitude and automatized explicit knowledge. However, as proficiency increases, the slope becomes steeper, signalling a stronger relationship. The same pattern holds true for implicit and explicit knowledge, though it is less pronounced. Overall, the observable pattern points towards greater effects of implicit aptitude at more advanced proficiency levels.

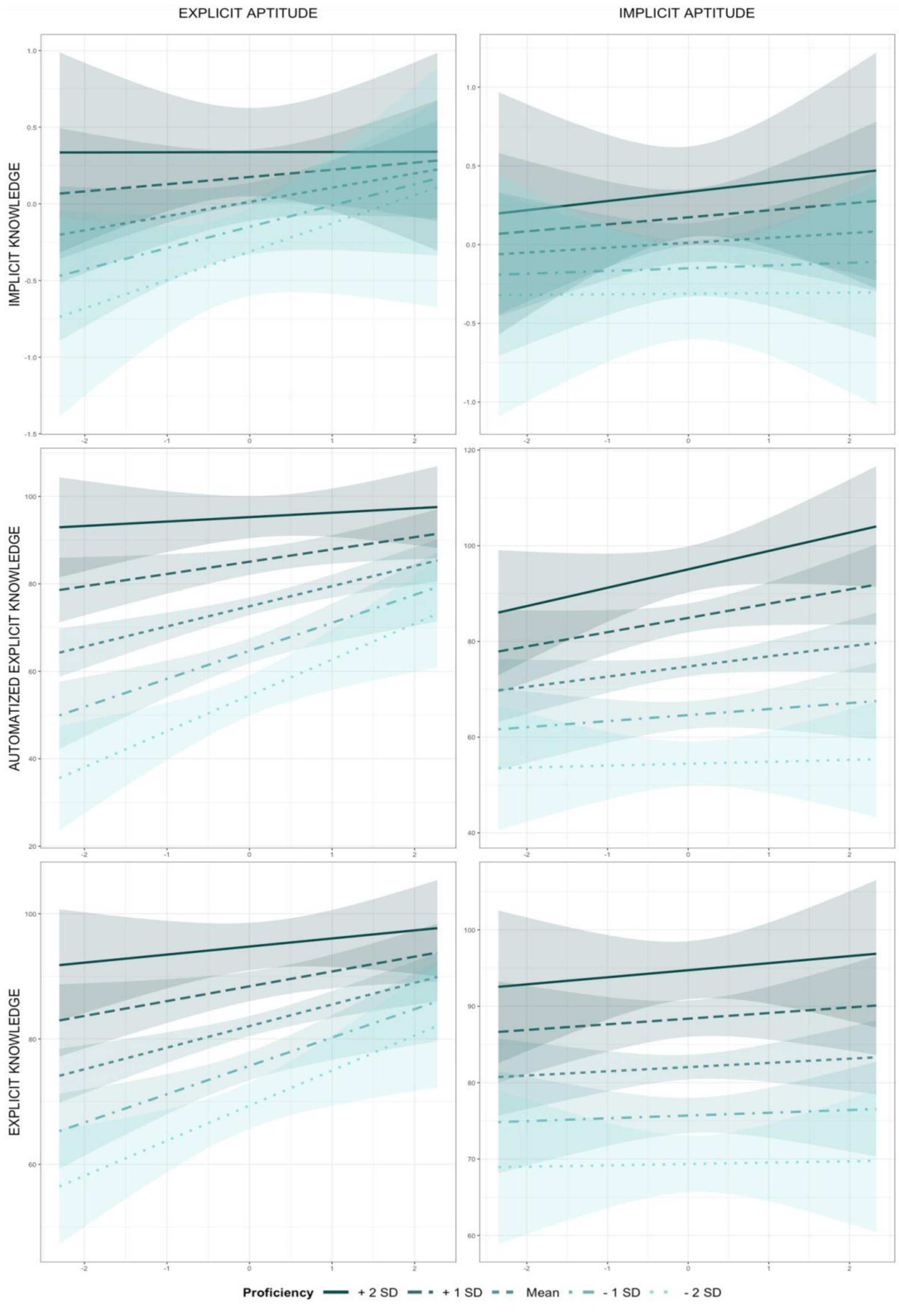

Figure 25. Relationships between L2 knowledge and cognitive aptitudes

Despite the absence of significant interaction terms in Model 3, which showed a significant negative main effect of WM on explicit knowledge, we plotted the relationship at five proficiency levels, as above. Figure 26 shows that, again, the slope of the relationship is different at different proficiency levels, pointing towards a (non-significant) interaction. Specifically, the observable pattern indicates that the negative effects of WM on explicit knowledge decrease as proficiency increases.

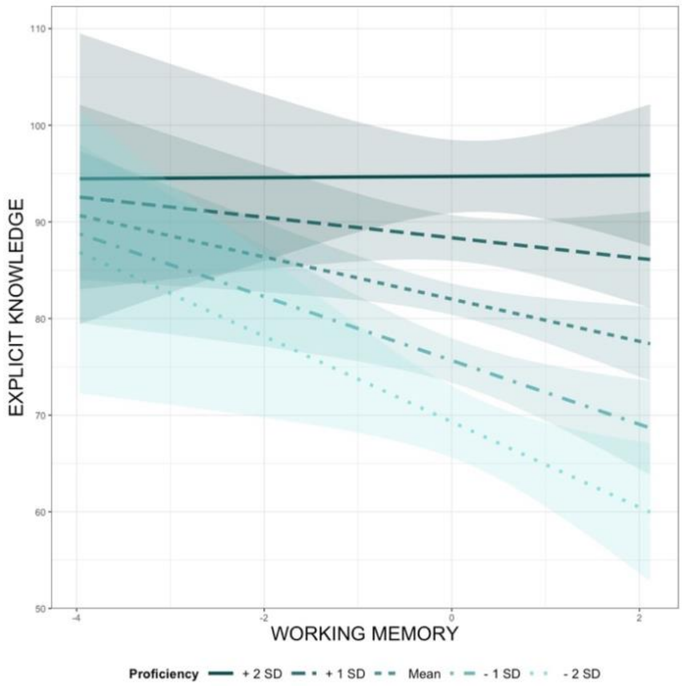

Figure 26. Relationship between explicit knowledge and working memory

5.6.3. At which point on the L2 proficiency continuum do the facilitative effects of explicit and implicit aptitude and working memory become or cease to be significant?

To address the second research question, we further scrutinized the interactions described in the preceding section by means of simple slopes analysis and Johnson-Neyman (J-N) intervals. Simple slopes analysis involves the methodical examination of the slope of a relationship between two variables at different points of a mediating variable. If and where the slope is significant, the relationship is significant. In our case, this means examining the slope of the relationship between aptitude and WM on the one hand and the L2 outcome measures on the other hand at different proficiency levels. We used the same points on the continuum as above, that is M-2SD, M-1SD, M, M+1SD, and M+2SD.

First, we focused on the significant interaction of proficiency with the relationship between explicit aptitude and automatized explicit knowledge as identified in Model 2. The results of the simple slopes analysis are shown in Table 17.

Table 17. Simple slopes analysis of proficiency as a mediator of the relationship between explicit aptitude and explicit knowledge

Proficiency level	Est.	Std. error	t	p
M-2SD	8.20	2.47	3.32	.000
M-1SD	6.41	1.61	3.98	.000
M	4.61	1.05	4.37	.000
M+1SD	2.81	1.29	2.18	.030
M+2SD	1.01	2.06	0.49	.620

As Table 17 indicates, the critical shift happens between M+1SD (p = .030) and M+2SD (p = .680) proficiency. For values of M+1SD and lower, the relationship is significant, while for values of M+2SD and higher, it is no longer significant.

J-N intervals capture the entire range of significance and non-significance of a mediating variable, so can help determine the precise point of change. A significant J-N interval represents a range of values of the mediating variable for which the slope of the relationship between the other two variables is significantly different from zero. A significant slope means a significant relationship, so the J-N intervals in fact show the range of proficiency where there is a significant relationship between the predictor and the outcome. Figure 10 shows the relevant J-N interval plot.

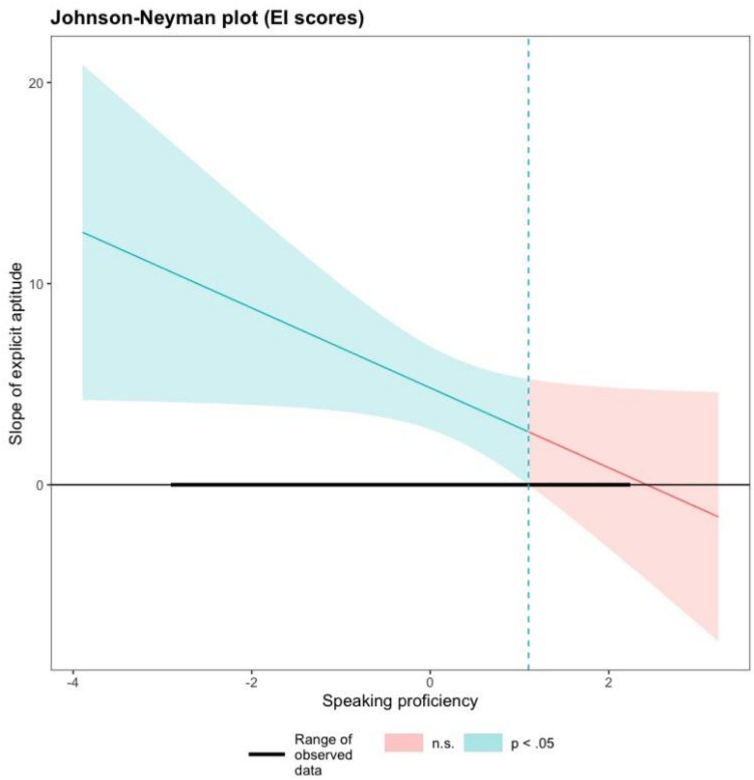

Figure 27. Johnson-Neyman plot of the interaction between speaking proficiency and relationship between explicit aptitude and EI scores

As can be seen in Figure 27, the slope coefficient as represented by the diagonal line is positive (above zero) and significant (turquoise) at lower levels of proficiency. As proficiency increases, the value of the slope decreases until it reaches non-significant levels (red). The simple slopes analysis further yields the interval of all observed (standardized) values [-2.89, 2.22] along with the interval of significance [-2.89, 1.10]. Taken together, this indicates that the relationship is significant for values of proficiency between M-2.89SD and M+1.10SD, while it is non-significant for values of proficiency between M+1.10SD and M+2.22SD. We can therefore conclude that the tipping point is M+1.10SD. Below this value, there is a significant positive relationship between explicit aptitude and automatized explicit knowledge. As proficiency increases, the magnitude of this relationship decreases, and from M+1.10SD proficiency onwards, the relationship is no longer significant.

Second, we examined the (non-significant) interaction of proficiency with the relationship between explicit aptitude and explicit knowledge (Model 3). The interval of all observed (standardized) values [-2.89, 2.22] remains the same. A simple slopes analysis (see Appendix 6 for full details) and calculation of the J-N intervals placed the critical point at M+1.13SD proficiency, as illustrated in Figure 28.

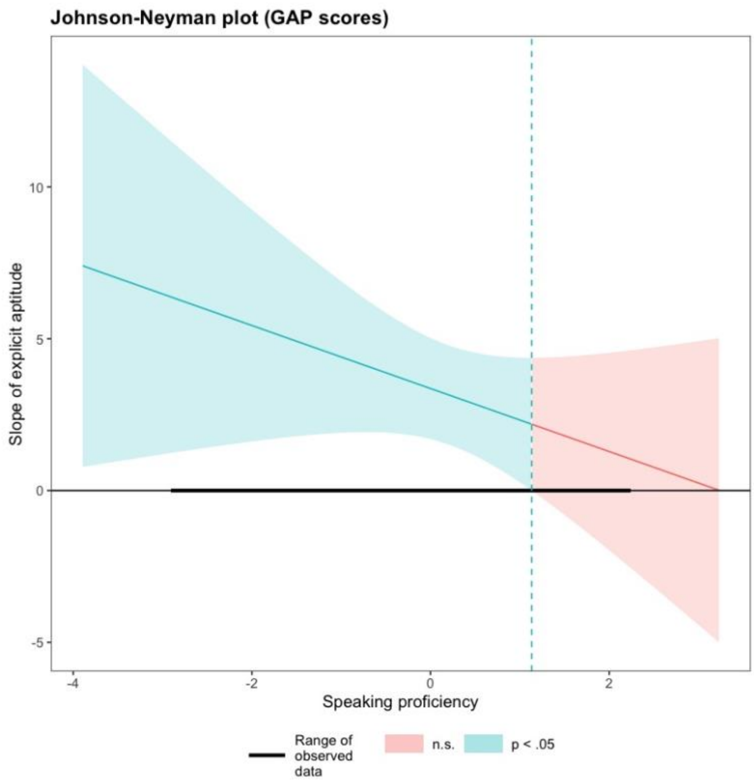

Figure 28. Johnson-Neyman plot of the interaction between speaking proficiency and relationship between explicit aptitude and GAP scores

The critical point at M+1.13SD for explicit knowledge is almost identical to the one detected for automatized explicit knowledge at M+1.10SD, indicating a consistent trend in the relationship between explicit aptitude and (automatized) explicit knowledge: as proficiency increases, the facilitative effect of explicit aptitude decreases.

Third, we focused on the (non-significant) interaction of proficiency with the relationship between implicit aptitude and automatized explicit knowledge (Model 2), again performing a simple slopes analysis (see Appendix 6 for full details) and examining the J-N intervals shown in Figure 29.

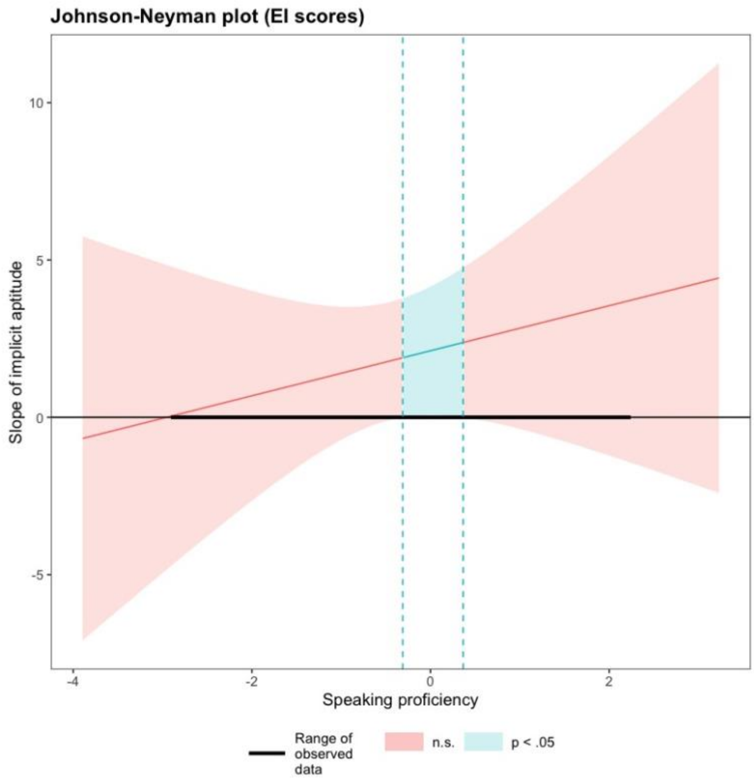

Figure 29. Johnson-Neyman plot of the interaction between L2 speaking proficiency and relationship between implicit aptitude and EI scores

Figure 29 suggests an overall positive relationship between implicit aptitude and automatized explicit knowledge, even though it is statistically non-significant except within a narrow range around the proficiency mean. Importantly, the strength of this relationship increases as proficiency increases, as indicated by the increasing values of the slope. This points towards a tipping point around mean proficiency after which implicit aptitude has an increasingly facilitative effect.

Fourth, we focused on the (non-significant) interaction of proficiency with the relationship between working memory and explicit knowledge (Model 3). We performed a final simple slopes analysis (see Appendix 6 for full details) and examined the J-N intervals shown in Figure 30.

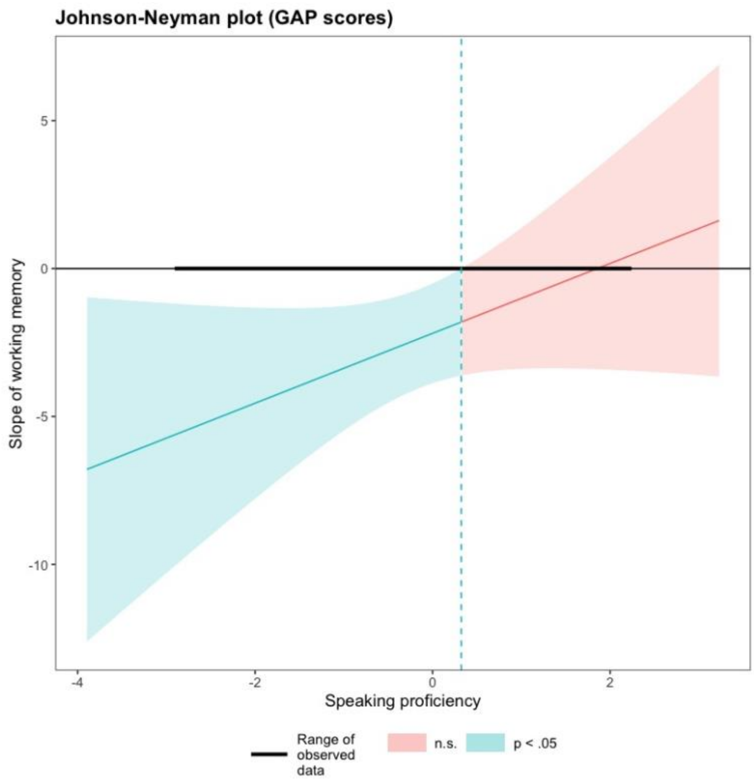

Figure 30. Johnson-Neyman plot of the interaction between L2 speaking proficiency and relationship between working memory and GAP scores

As can be seen in Figure 30, the critical point is very close to the mean at M+0.32SD proficiency. This suggests that the negative effect of WM on explicit knowledge stops being significant around mean proficiency.

Finally, we examined all other relationships and potential significant interactions presented in Figure 24. However, as illustrated in Figure 24 and detailed in Appendix 5, no additional relationships or interactions reached significance. Although some visual patterns in Figure 25 suggest a significant interaction – specifically between proficiency and the relationship between explicit aptitude and implicit knowledge – no such interaction was found. Consequently, no additional J-N plots were produced.

5.7. Discussion

The current study aimed to determine the mediating influence of L2 proficiency in the relationship between explicit and implicit aptitude and WM on the one hand and L2 outcomes in terms of explicit, automatized explicit and implicit knowledge of selected morphosyntactic structures on the other hand. To this end, we first established how the cognitive ability variables factored together. We arrived at three factors which were in line with recent findings (Granena, 2012, 2013a; Roehr-Brackin et al., 2023): explicit aptitude (LLAMA B, E, F), implicit aptitude (LLAMA D and SRT), and working memory (FDS and OSPAN). Using multilevel modelling, we examined main effects and interactions, with the former revealing the predictive power of the cognitive ability factors and the latter the mediating influence of L2 proficiency. On the basis of simple slopes analysis and J-N intervals, we identified the tipping points at which the mediating influence of proficiency underwent a critical change, either losing or gaining facilitative impact.

5.7.1. The role of working memory at different proficiency levels

Contrary to expectation, we found a significant negative main effect of WM in the model predicting explicit knowledge, that is, higher levels of WM were associated with lower scores on the gap-fill task used in the present study. A (non-significant) interaction with proficiency indicated that the negative relationship was restricted to lower proficiency levels. While seemingly counter-intuitive, our finding is reminiscent of the role of switching ability reported by Linck et al. (2013), who identified a negative effect of this executive function. While an explanation is not immediately obvious, it is possible that heavy reliance on WM and executive function in particular may counteract the use of explicit knowledge in contexts where there is no time pressure and where the use of metalinguistic knowledge is encouraged. In such a scenario, reliance on explicit aptitude rather than WM would appear to be advantageous, at least at lower proficiency levels. This is indeed what our study showed.

5.7.2. The role of explicit aptitude at different proficiency levels

Explicit aptitude emerged as a significant positive predictor of automatized explicit and explicit knowledge – a finding which further substantiates previously reported associations between explicit aptitude and outcome measures targeting or favouring the use of explicit knowledge (Li, 2015, 2016). Importantly, we found that proficiency was a mediator in this relationship: the role of explicit aptitude decreased with increasing proficiency, significantly so for automatized explicit knowledge measured via an elicited imitation task, and observable as a trend for explicit knowledge measured via a gap-fill task. Put differently, explicit aptitude is of benefit at lower levels of proficiency, but does not convey any additional advantages at higher levels of proficiency.

Our finding aligns with previous research on the role of explicit aptitude in beginner- and intermediate-level learners (Artieda & Muñoz, 2016; Li, 2015, 2016) while also providing additional empirical evidence that explicit aptitude has a limited impact at higher levels of L2 proficiency (Morgan-Short et al., 2014). This supports the argument that explicit aptitude is crucial at initial stages (Carroll, 1990) when learners are likely to face challenges, in particular with regard to morphosyntax, as targeted in our study as well as a number of previous studies. For instance, Li (2013) found that language-analytic ability as a component of explicit aptitude played a crucial role when learners were confronted with a hard and opaque structure; Yalçın and Spada (2016) reported that language-analytic ability was relevant for judging the grammaticality of difficult structures.

Beyond corroborating previous research, our findings represent a critical step forward for current research, since we were able to identify the exact point at which facilitative effects of certain aptitude components ceased or came into play, as appropriate. In the case of explicit aptitude predicting automatized explicit knowledge, we can specify the scores that would need to be achieved on the components of our speaking proficiency measure for the effects of explicit aptitude to become non-significant. To illustrate, we identified M+1.10SD as the tipping point (see Section 5.6). As speaking proficiency was operationalized in terms of standardized factor scores, we can interpret the threshold by looking for participants who scored at or near the relevant value and then trace back their performance on the component measures. The nearest scores to M+1.10SD in our data set are standardized proficiency scores of 1.09 and 1.11. These correspond to the following combination of speaking proficiency values: Giraud's index (lexical complexity) between 7.65 and 8.61, ratio of clauses to utterances (morphosyntactic complexity) between 1.65 and 2.10, lexical and morphosyntactic errors per 100 words of 0.00-0.47 and 0.00-0.23, respectively, and fluency of 105-127 words per minute. In other words, in terms of speaking proficiency components, these are the points at which explicit aptitude ceases to have beneficial effects for performance on an elicited imitation task aimed at assessing automatized explicit knowledge.

It goes without saying that the above interpretation is only meaningful in the context of the present study, since it is dependent on our specific operationalizations. We have, however, presented a proof of concept which is clearly generalisable beyond the current study: it is possible to identify an exact threshold on the proficiency continuum at which specific cognitive abilities (cease to) come into play.

5.7.3. The role of implicit aptitude at different proficiency levels

Implicit aptitude was a significant positive predictor of automatized explicit knowledge. This effect was mediated (non-significantly) by proficiency. The overall trend showed that the facilitative effects of implicit aptitude were absent at lower levels and became more pronounced as proficiency increased. In other words, higher implicit aptitude benefitted learners at higher proficiency levels. This finding aligns with previous research which has emphasized the importance of implicit aptitude for high achievers (Linck et al., 2013) or in later stages of learning (Morgan-Short et al., 2014). The implications are twofold: at lower proficiency levels, learners with low implicit aptitude are not disadvantaged, nor do learners with high implicit aptitude derive benefit. However, at higher proficiency levels, learners with high implicit aptitude can leverage it to their advantage, while those with low implicit aptitude may be hindered.

Interestingly, implicit aptitude was not a significant predictor of implicit knowledge and one possible explanation is learning context. Suzuki and DeKeyser (2015) found that the relationship between implicit aptitude and implicit knowledge was present only in learners with extended lengths of residence in a target-language-speaking country. In contrast, the vast majority of participants in the current study had not spent much or any time in an immersion environment. Of course, this interpretation remains speculative at this stage.

5.8. Conclusion

The current study has confirmed the predictive power of explicit and implicit aptitude and WM in (instructed) L2 learning. Moreover, the study offers further evidence for the mediating role of L2 proficiency in the relationship between cognitive ability factors and L2 outcomes. In sum, our findings show that as proficiency increases, the beneficial influence of explicit aptitude diminishes while a facilitative influence of implicit aptitude kicks in. Finally, and arguably most importantly, we have demonstrated that it is possible to pinpoint the threshold on the proficiency continuum at which a critical switch away from or towards significant facilitation occurs.

5.9. Limitations and suggestions for future research

Inevitably, the present study had a number of limitations, three of which we will mention here. First, our outcome measures focused on selected morphosyntactic structures of English, offering insight into a limited domain of achievement in a specific L2. Second, our proficiency measure, though multi-componential in nature, focused on speaking skill, whereas a fully comprehensive measure of proficiency would include listening, reading and writing as well. Third, the interpretation of the significant tipping point which we identified for the influence of explicit aptitude on automatized explicit knowledge is only meaningful in the context of the current study.

Future research should aim to address the above limitations by including different L2s and different linguistic focus points in their outcome measures and by considering a more comprehensive proficiency measure. In order to achieve a level of generalisability that can be of practical use, proficiency scores benchmarked against the Common European Framework of Reference or other internationally recognised standards would be most desirable. At the same time, we believe that the theoretical and empirical contribution of our findings to the field of L2 learning research is readily apparent and that wider use of the analytic approach employed in our study offers a new and fruitful avenue for future research.

6. Chapter 6 (Article 3) – At the interface of language aptitude and L2 knowledge

6.1. Abstract

The relationship between implicit and explicit knowledge has been widely debated, yet empirical studies directly examining this remain limited. This study addresses this gap by examining the relationship between cognitive aptitudes and L2 knowledge while acknowledging the mediating role of proficiency and adopting a contemporary view of explicit and implicit aptitude as multi-componential constructs. Eighty-three Croatian learners of English were assessed on implicit and explicit aptitude using the LLAMA suite and an SRT task. Implicit knowledge was measured via a self-paced reading task, automatized explicit knowledge through an elicited imitation test, and explicit knowledge with a gap-fill test. Multilevel modeling that accounted for individual differences and included proficiency as a mediator revealed that explicit aptitude significantly predicted implicit knowledge, supporting the explicit-implicit interface. Additionally, implicit aptitude significantly predicted automatized explicit knowledge, revealing an implicit-explicit relationship. These findings suggest a bidirectional and dynamic interface between implicit and explicit learning and knowledge.

6.2. Introduction

Most L2 learners recognize that beyond the linguistic rules they consciously learn and can recall through introspection, there are aspects of language that operate outside their conscious awareness yet are nevertheless executed correctly (Bialystok, 1994b). In an effort to address this duality, SLA researchers have proposed distinct representations for these two facets of knowledge: implicit and explicit. However, more than four decades after Krashen's monitor model (1981, 1982, 1985), which initially proposed a distinction between these two types of L2 knowledge, the question of whether and how they interact over time remains unresolved.

Awareness has frequently been a construct in focus since it is the criterion by which implicit can be distinguished from explicit knowledge. The question, however, has not been merely the presence of awareness, but rather the extent of the importance it was granted. Perspectives on this matter range from Krashen's (1981, 1982, 1985) view that awareness has minimal impact on L2 acquisition, to arguments positing it as a critical step in L2 learning (Bialystok, 1994a; DeKeyser, 2015, 2020; Schmidt, 2001, 2012), with various intermediate positions (N. Ellis, 2005, 2015; R. Ellis, 1994, 2005). Recent empirical investigations have explored this issue across laboratory (Andringa, 2020; Curcic et al., 2019) and naturalistic setting (Kim & Godfroid, 2023; Suzuki & DeKeyser, 2017) consistently underscoring both the significance of awareness in L2 acquisition and the relationship between implicit and explicit knowledge. However, the designs of these studies varied considerably. While a longitudinal approach enables direct examination of the relationship between the two types of knowledge, it often suffers from high attrition rates, small sample sizes, and limited statistical power. Similarly, laboratory studies suffer from small sample sizes and low ecological validity. In contrast, Suzuki and DeKeyser (2017) study demonstrated that a cross-sectional design can be a viable alternative to longitudinal studies for investigating the relationship, provided that both the cognitive abilities underlying the learning process and the resulting linguistic knowledge are measured in order to infer causality.

The present study seeks to deepen our understanding of the relationship between explicit and implicit knowledge by adopting a cross-sectional design to analyse the interplay between cognitive abilities and L2 knowledge. A comprehensive range of measures of L2 knowledge and underlying cognitive abilities was employed, acknowledging the multifaceted nature of language aptitude. Additionally, a range of proficiency levels was included to address potential mediating effects. Factor analysis was utilized to identify measures of implicit and explicit aptitude and knowledge, while a multilevel modeling approach was employed as a robust statistical technique capable of accounting for individual learner differences (Linck, 2016). Crucially, predictive relationships between aptitude for implicit and explicit learning and implicit, explicit, and automatized explicit knowledge were scrutinized as a way of establishing explicit-implicit and implicit-explicit relationships.

6.3. Background

6.3.1. Implicit and explicit learning and knowledge

Implicit learning involves the acquisition of unconscious knowledge without conscious intent or awareness. This type of knowledge is characterized as tacit, intuitive, and beyond conscious introspection, typically allowing for rapid access (Hulstijn, 2015; Rebuschat, 2013). In contrast, explicit learning involves the acquisition of conscious, declarative knowledge through intentional learning processes (Rebuschat, 2013). Unlike implicit knowledge, explicit knowledge is conscious, available for introspection, and potentially verbalizable (N. Ellis, 2015; Rebuschat, 2013). Awareness, defined as the conscious perception of what is being learned (Williams, 2009), plays a critical role in differentiating implicit from explicit learning and knowledge. The distinction between these two types of knowledge also extends to the underlying memory systems: implicit knowledge is supported by procedural memory, while explicit knowledge is supported by declarative memory (Paradis, 2004, 2009). Although the declarative/procedural distinction is not strictly equivalent to implicit/explicit from the perspective of cognitive neuroscience, in SLA these terms are often used synonymously for practical purposes (DeKeyser, 2017). Different types of knowledge emerge from distinct processing activities at various stages of the acquisition process (Leow, 2015). Consequently, a focus on meaning may lead learners to develop different knowledge than a focus on form (see (see Godfroid, 2023, for a concise overview). Understanding the relationship between these two types of knowledge is essential for gaining deeper insights into the role of awareness in L2 learning. This relationship is often referred to as the "interface issue" and has been a point of contention in the field of SLA since its conception more than forty years ago.

6.3.2. Measures of implicit and explicit knowledge

R. Ellis's (2005) seminal work provides a framework for measuring implicit and explicit knowledge. In tests designed to assess implicit knowledge, learners are expected to respond based on intuition with a high degree of certainty, and their responses should exhibit consistency. The primary focus of these tasks should be on meaning, with time pressure applied, and the targeted language features should preferably be early-introduced language features. Importantly, learners are not required to possess any metalinguistic knowledge for completing these tasks. Conversely, tests aimed at assessing explicit knowledge should elicit responses based on systematicities in the language. These tasks impose no time pressure and focus on linguistic form, typically encouraging the use of metalinguistic knowledge. Within this framework, oral narrative, elicited imitation (EI), and timed written grammaticality judgment tests are suggested as measures of implicit knowledge, while untimed written grammaticality judgment tests and test of metalinguistic knowledge are proposed as measures of explicit knowledge.

Subsequently, Godfroid and Kim (2021) examined the validity of a battery of nine tests, ranging from more explicit, such as metalinguistic knowledge and untimed grammaticality judgment tests, to more implicit, such as elicited imitation, self-paced reading (SPR), and word monitoring. Their factor-analytic results demonstrated a lack of predictive relationship between implicit sequence learning ability, measured by a serial reaction time (SRT) task, and reaction-time measures like self-paced reading and word monitoring, casting doubt on their validity as indicators of implicit knowledge, potentially due to their relatively low reliability in the study. However, the significant predictive relationship detected between implicit aptitude and both the timed grammaticality judgment test and elicited imitation test indicated that timed, accuracy-based tests were valid measures of implicit knowledge, corroborating earlier findings by R. Ellis (2005).

However, in another study, elicited imitation was found to measure a construct different from implicit knowledge (Suzuki & DeKeyser, 2015). Regression analysis revealed a lack of a predictive relationship between implicit aptitude, assessed by a serial reaction time task, and elicited imitation. A significant bivariate relationship between elicited imitation and metalinguistic knowledge scores suggested that the elicited imitation task may actually tap into explicit knowledge. Given that elicited imitation is an online task requiring oral production, the knowledge accessed through this task must be retrieved quickly, leading researchers to conclude that elicited imitation measures automatized explicit knowledge (Suzuki & DeKeyser, 2015). This type of knowledge is defined as conscious (explicit) knowledge that has undergone varying levels of automatization. Both Suzuki (2017) and Vafaee et al. (2017) argue that elicited imitation measures automatized explicit knowledge, with recent neurolinguistic evidence indicating reliance on declarative memory during the production phase of the task (Suzuki et al., 2023).

These findings cast doubt on R. Ellis's (2005) framework, particularly the criterion of time pressure. Suzuki and DeKeyser (2015) argue that time pressure alone is insufficient to limit access to explicit knowledge if such knowledge has been automatized and can be accessed rapidly. Since implicit and explicit knowledge can be distinguished based on the criterion of awareness, the authors suggest that measures of implicit knowledge should limit learners' awareness by assessing the knowledge at the processing level. Reaction time-based measures such as word monitoring (Godfroid, 2016; Granena, 2013b; Kilborn & Moss, 1996; Marslen-Wilson & Tyler, 1980) and self-paced reading (Jiang, 2007; Jiang et al., 2011; Roberts & Liszka, 2013) are recommended over accuracy-based measures like elicited imitation and timed grammaticality judgment tests. This is because reaction time measures that assess sensitivity to grammatical violations are presumed to operate outside participants' awareness, making them "pure" measures of implicit knowledge.

Similarly, Cleeremans's radical plasticity thesis (2008, 2011) suggests that implicit knowledge is characterized by shallow processing and weak memory representations. Within this view, implicit knowledge can only be detected using associative tasks that gauge knowledge at the comprehension, rather than productive, stage. This reinforces the idea that reaction-time measures have an advantage over accuracy-based measures in gauging implicit knowledge.

6.3.3. The interface debate

The interface issue refers to the potential theoretical associations and mutual influence between implicit and explicit knowledge in L2 learning.

6.3.3.1. No Interface Position

The no-interface position asserts that explicit and implicit knowledge are acquired through distinct mechanisms, with no relationship between them (Krashen, 1981). According to this view, there is no facilitation or interaction between these two types of knowledge. Krashen further argues that language is too complex to be learned explicitly, thereby relegating the role of explicit knowledge to merely monitoring L2 performance. This position has been highly controversial since its inception (Han & Finneran, 2014), largely due to substantial empirical evidence suggesting that explicit knowledge can positively contribute to implicit knowledge development (R. Ellis, 2002; Norris & Ortega, 2000; Russell & Spada, 2006). While Paradis

(2004, 2009) and Hulstijn (2015) maintain a theoretically similar stance, they propose an indirect influence of explicit knowledge on implicit learning, wherein explicit knowledge guides learners' practice, thereby indirectly supporting the development of implicit knowledge.

6.3.3.2. Interface Position

The strong interface position is rooted in skill acquisition theory (DeKeyser, 2015, 2020) which posits that language learning progresses through distinct stages: from declarative ("knowledge that") to procedural ("knowledge how"), and finally to automatic (spontaneous and effortless) knowledge. According to this view, implicit knowledge is derived from explicit knowledge through repeated practice (DeKeyser, 1998, 2020; Sharwood Smith, 1981). In more recent elaborations, DeKeyser (2017, 2020) clarifies that explicit knowledge does not directly transform into implicit knowledge but rather plays a causal role, a claim also supported by recent empirical findings (Andringa & Curcic, 2015; Zhang, 2015).

In contrast to the strong interface position, the weak interface position is more nuanced and encompasses multiple views. One perspective suggests that learners can use their explicit knowledge to generate output which subsequently serves as self-generated input for implicit learning mechanisms (Schmidt & Frota, 1986; Sharwood Smith, 1981). R. Ellis (1994; 2005, 2006) contends that implicit knowledge can be derived from explicit knowledge, but this depends on the nature of the linguistic elements involved. N. Ellis (2005, 2006; 2015) argues that learners acquire implicit knowledge through probabilistic encounters with relevant exemplars in their environment, gradually organizing this knowledge over time. However, because this process is imperfect, explicit instruction, feedback, and knowledge can help, facilitating the initial registration of patterns that are then refined and integrated through implicit learning.

6.3.3.3. Implicit-explicit Interface Position

Another perspective on the interface issue emerges from developmental and cognitive psychology, where research suggests that rule awareness (i.e. explicit knowledge) can develop from implicitly acquired knowledge. According to Bialystok (1994a, 1994b, 2001), learners begin with unanalysed knowledge, and language learning is essentially a process of increasing explicitness, or developing awareness of linguistic structures through the process of analysis. Evidence in favour comes from studies showing that accuracy rates on tasks following some form of implicit learning are significantly lower for participants that remained unaware of the underlying system being learned, compared to those who developed awareness in the process (Andringa, 2020; Williams, 2005). This line of research suggests that implicit knowledge can inform or contribute to explicit knowledge, providing a basis for the implicit-explicit interface (Godfroid, 2023).

Cleeremans (2008, 2011) offers another perspective through his radical plasticity thesis, which posits that knowledge develops on a continuum across three stages. It begins as implicit knowledge, which is unconscious and characterized by low depth of processing. With increasing exposure, the strength and quality of associations improve, making this knowledge accessible for conscious introspection thus forming explicit knowledge. Finally, with continued exposure, the knowledge becomes so strong that it is no longer subject to conscious control, instead influencing behaviour automatically. At this stage, it is referred to as automatized explicit knowledge, representing the endpoint of the developmental continuum. According to Godfroid (2023), the interface exists at the point of transition between stages one and two. The shift from the implicit to explicit stage (from stage one to stage two) is described as representational redescription, a concept introduced by Karmiloff-Smith (1992). This process involves a general, cyclical recoding of information from a less to a more compressed format, which Bialystok (1994a, 1994b, 2001) refers to as the process of analysis.

6.3.3.4. Co-existence of Interfaces

In another development of the ongoing debate, Han and Finneran (2014) propose the simultaneous existence of multiple interfaces between explicit and implicit knowledge, which do not align with any single interface position. Instead, they argue that different relationships exist within and across various linguistic subsystems, as well as in different L2 learners. They contend that certain structures, particularly those at the interface between syntax and semantics, syntax and pragmatics, and syntax and discourse, are prone to a permanent lack of interface. Additionally, linguistic features that are often used correctly on tasks favouring explicit knowledge, but not on those favouring implicit knowledge, such as articles or inflectional morphemes, are also susceptible to a permanent lack of interface. In other words, different aspects of grammar exhibit varying degrees of susceptibility to strong, weak, and no interface relations.

6.3.4. Empirical investigations into the interface issue

6.3.4.1. Longitudinal approach

Kim and Godfroid (2023) conducted a longitudinal study to examine the relationship between implicit and explicit knowledge in L2 learners of English. Utilizing R. Ellis's (2005) measurement framework, the study assessed implicit knowledge through timed written grammaticality judgment tests (GJT), elicited imitation, and oral production, while explicit knowledge was measured using untimed grammaticality judgment tests (GJT) and metalinguistic knowledge (MLK) tests. Data were collected at two time points to enable causal inferences about the relationship between the types of knowledge. The researchers compared a non-interface model, an interface model (explicit-implicit interface), and a reciprocal interface model (explicit-implicit and implicit-explicit interface).

The results indicated that the best fit was found in the reciprocal interface model, suggesting a bidirectional relationship: prior explicit knowledge influences the development of implicit knowledge, and implicit knowledge, in turn, affects the development of explicit knowledge. The former supports the concept of the explicit-implicit interface, where explicit knowledge contributes to the formation of implicit knowledge, aligning with N. Ellis's (2005, 2006; 2015), R. Ellis's (1994; 2005, 2006), and DeKeyser's (1998, 2015, 2020) theoretical perspectives. Notably, the study also revealed that explicit knowledge is significantly shaped by prior levels of implicit knowledge, indicating an implicit-explicit interface. The authors proposed that implicit knowledge might play a role in the discovery of explicit rules, as suggested by Bialystok (1994b, 2001) and Cleeremans (2008). In conclusion, the study offers an interesting perspective on the interface issue, viewing explicit and implicit knowledge as dynamically interacting throughout the learning process.

While the study's longitudinal approach is commendable, the constructs of implicit and explicit knowledge, derived from factor analysis, exhibited a strong intercorrelation ($r = .83$), potentially due to the effects of repeated testing, which raises concerns about their discriminant validity. The use of elicited imitation and grammaticality judgment tasks to measure implicit knowledge is arguably questionable, as these tasks have previously been shown to align more closely with automatized explicit rather than implicit knowledge (Suzuki, 2017; Suzuki & DeKeyser, 2015). This calls into question the validity of the study's latent factors of implicit and explicit knowledge, and consequently, suggests potential concerns regarding the existence of the interfaces detected in the study.

6.3.4.2. Cross-sectional approach

Suzuki and DeKeyser (2017) conducted a cross-sectional study to investigate the interface issue, specifically examining whether automatized explicit knowledge contributes to the acquisition of implicit knowledge. The study utilized word monitoring, self-paced reading, and visual-world tasks to assess implicit knowledge, while timed auditory and written grammaticality judgment tests, along with timed fill-in-the-blank tests, were employed to evaluate automatized explicit knowledge in advanced L2 learners of Japanese. Aptitude for explicit learning was defined as language-analytic ability, measured using the LLAMA F subtest, while aptitude for implicit learning was defined as implicit sequence learning ability, assessed through a serial reaction time task.

The researchers compared two structural equation models: a non-interface model, where implicit and automatized explicit knowledge are influenced by cognitive aptitudes but remain unrelated, and an interface model which included a path from automatized explicit knowledge to implicit knowledge. Although neither model was statistically superior, the positive and significant path from automatized explicit to implicit knowledge ($r = .35$) in the interface model suggested an interaction between the two types of knowledge. This finding supports the idea that automatized explicit knowledge can facilitate the development of implicit knowledge. The researchers hypothesized that this facilitation could occur in two ways. First, automatized explicit knowledge may help learners process language input more efficiently by directing their attention to relevant grammatical features, which allows the implicit learning system to gradually pick up these features. Second, automatized explicit knowledge may enable the frequent and accurate use of relevant grammatical structures, which then serves as input feeding into implicit learning system.

In conclusion, the study provided indirect evidence of the explicit-implicit interface through a cross-sectional approach by focusing on the relationship between underlying cognitive abilities (explicit and implicit aptitude) and different types of L2 knowledge. This offers a novel perspective on the interface issue, examining it through the lens of the relationship between cognitive aptitudes and L2 knowledge.

However, Suzuki and DeKeyser (2017) did not include a measure of less automatized explicit knowledge, leaving open the question of whether the facilitative role of automatized explicit knowledge in the development of implicit knowledge extends to explicit knowledge that is not automatized. Instead, the study suggests that explicit knowledge undergoes a process of automatization, which then relates to implicit knowledge – a claim that assumes a unidirectional interface. In contrast, the findings of Kim and Godfroid (2023) suggest a dynamic interplay of explicit and implicit knowledge in both directions, a theoretical notion that cannot be fully tested without measuring non-automatized explicit knowledge. Additionally, Suzuki & DeKeyser's study (2017) did not find direct evidence for the influence of explicit aptitude on implicit knowledge, revealing only an indirect effect – aptitude for explicit learning influences automatized explicit knowledge, which is then related to implicit knowledge. This explanation confines the results to the framework of skill acquisition theory (DeKeyser, 2015, 2020), where the interface is seen as existing only in the explicit-implicit direction. Furthermore, the study assessed explicit aptitude using language-analytic ability and implicit aptitude using implicit sequence learning ability, despite contemporary views suggesting that both explicit and implicit aptitude are multi-componential (Godfroid & Kim, 2021; Granena, 2012, 2013a, 2020; Li & DeKeyser, 2021; Roehr-Brackin et al., 2023) (Pavlekovic & Roehr-Brackin, 2024) [Chapter 4]. Finally, studies such as Artieda and Muñoz (2016) and Pavlekovic and Roehr-Brackin (under review) [Chapter 5] indicate that L2 proficiency mediates the relationship between language aptitude and L2 knowledge, so the inclusion of only advanced-level learners might have skewed the results, a point that the authors themselves acknowledge.

6.4. The current study

In summary, existing research has demonstrated that both accuracy-based and reaction-time-based measures can effectively assess L2 knowledge, but agreement has yet to be reached regarding which measures specifically capture implicit versus (automatized) explicit knowledge. The investigation of the aptitude–knowledge relationship has been validated as a viable alternative to longitudinal designs for exploring the interface issue. However, previous studies have been limited by a narrow proficiency range and a restricted set of measures for cognitive aptitudes. Consequently, they have failed to account for the mediating role of proficiency in the relationship between cognitive aptitudes and L2 outcomes, and have been unable to find direct evidence of an interface. While the implicit-explicit interface has theoretical support, it has only been observed in a single out-of-laboratory study that utilized factor scores with poor discriminant validity. In light of these limitations, the present study posed the following research questions:

RQ6: How are scores on self-paced reading, elicited imitation, and gap-fill tests related?

RQ7: How do aptitude for explicit learning and aptitude for implicit learning contribute to learners' implicit knowledge?

RQ8: How do aptitude for explicit learning and aptitude for implicit learning contribute to learners' explicit and automatized explicit knowledge?

6.5. Methodology

The following sections provide a comprehensive overview of the participant sample, instruments, and procedures of the present study. A broad array of measures was employed to assess cognitive factors, including online administration of the LLAMA test suite and a probabilistic serial reaction time task. Linguistic knowledge of specific grammatical structures was evaluated using a self-paced reading task as a measure of implicit knowledge, while automatized explicit knowledge was assessed through an elicited imitation test. Explicit knowledge was measured using a gap-fill test (GAP). Lastly, speaking proficiency was evaluated through a 3-minute oral production task, and reading and listening proficiency were assessed using the Oxford Placement Test.

6.5.1. Participants

The study sample consisted of 83 L1 Croatian speakers learning English, aged between 14 and 18 years (M = 16.16, SD = 1.28). The group included 59 females, 22 males, and 2 participants who chose not to disclose their gender. All participants were enrolled in secondary education and had studied English for 6 to 13 years (M = 10, SD = 1.69) as part of their compulsory school curriculum.

6.5.2. Instruments and Procedures

All measures, with the exception of the oral production task and Oxford Placement Test, were developed using PsychoPy and administered through the Pavlovia platform (Peirce et al., 2019). Participants were provided with instructions in their L1 and were required to use headphones in a quiet environment. The author supervised the sessions via Zoom to ensure protocol adherence and to address any questions.

6.5.3. Explicit and Implicit Aptitude

Language learning aptitude was assessed using the LLAMA test suite (Meara, 2005; Meara & Rogers, 2019) and a probabilistic serial reaction time task (Kaufman et al., 2010). The LLAMA suite comprises four subtests:

LLAMA B evaluates associative memory by requiring participants to learn 20 new vocabulary items associated with novel picture stimuli during a two-minute learning phase. The subsequent untimed testing phase involves matching the presented words to the corresponding pictures from a set of 20 options, with a maximum score of 20.

LLAMA D assesses auditory pattern recognition ability. Participants are exposed to 10 words in an unfamiliar language during the exposure phase. In the test phase, they listen to both familiar and new words and indicate their familiarity with a yes/no response. The maximum score is 40.

LLAMA E measures the ability to learn sound-symbol correspondences. Participants are presented with 24 phonetic symbols, each representing a distinct syllable, and have two minutes to learn the associations. In the untimed test phase, participants hear a combination of two syllables and must select the correct pair from 20 combinations of symbols previously encountered separately. The maximum score is 40.

LLAMA F is a grammatical inferencing task that gives participants four minutes to deduce the rules of an artificial mini-language. During the learning phase, participants interact with buttons revealing picture stimuli accompanied by written descriptions. The untimed test phase involves selecting word combinations that accurately describe the given pictures. The maximum score is 132.

The probabilistic SRT task involves reacting to visual stimuli presented as squares appearing in one of four horizontal locations on a computer screen. The sequence of stimuli follows a probabilistic rule, with 85% of the stimuli conforming to a predetermined training sequence and 15% following a control sequence. The task consists of 960 trials organized into 8 blocks of 120 trials each. Performance is measured by the difference in mean response times between the training and control conditions.

6.5.4. Target Structures and Linguistic Knowledge

Three grammatical structures were selected for this study: the past simple tense, the passive voice, and the use of articles. These structures were chosen based on the grammar syllabus used in the participants' English classes and common errors identified by their English teacher (D. Linić Učur, personal communication, 06/07/2020). Linguistic knowledge was assessed using three measures that range from more implicit to more explicit:

A self-paced reading task was used to assess implicit knowledge (Jiang, 2007; Jiang et al., 2011; Roberts & Liszka, 2013). During the task, participants read sentences word by word as quickly as possible in a moving-window presentation. They answered comprehension questions to ensure focus on meaning. The task included 72 target sentences, evenly split between grammatically correct and incorrect, and 24 filler sentences, all grammatically correct. Repetition of the same structure was avoided by a randomized order of sentences and by making sure there were no more than two consecutive sentences testing the same grammatical structure. Reaction times (RTs) were recorded at three regions of interest (ROIs): the critical word where a grammatical error could be detected, the word following the critical word, and the word following ROI 2. A delay in RT was expected at ROI 1 for ungrammatical sentences, with ROIs 2 and 3 capturing any spillover effects. A grammatical sensitivity index was calculated by subtracting RTs for grammatically correct sentences from RTs for ungrammatical sentences across all ROIs (Suzuki, 2017).

An elicited imitation test was used to measure automatized explicit knowledge (Suzuki & DeKeyser, 2015) during which participants listened to sentences followed by five random numbers. They were required to read each number on the screen sequentially and then repeat the sentence in correct English. Random numbers ensured active engagement with the numerical cues, preventing rote repetition of sentences (Mackey & Gass, 2022). The test encompassed 72 items across the three grammatical structures, evenly distributed between grammatically correct and incorrect sentences. Similar to the SPR task, repetition of the same structure was minimized to no more than two consecutive sentences. A nine-second time constraint was used for all sentences. The maximum score was 72.

A gap-fill test was utilized to measure explicit knowledge using a three-way multiple-choice format. Participants were given unlimited time. The task comprised 75 sentences, equally divided among the three targeted grammatical structures, with a maximum score of 75.

6.5.5. L2 Proficiency

Speaking proficiency was assessed through a 3-minute oral production task in which participants spoke about their typical day at school. Their speech was recorded, transcribed, and analysed using the CLAN software (MacWhinney, 2000). The analysis focused on complexity, accuracy, and fluency (CAF) measures, with lexical complexity assessed using Guiraud's Index (Guiraud, 1954) and morphosyntactic complexity based on the number of clauses per c-unit (Foster et al., 2000). Accuracy was calculated by counting errors per 100 words, and fluency was measured using pruned speech rate (words per minute)(Freed, 2000).

Reading and listening proficiency were assessed using the Oxford Placement Test(Oxford University Press, n.d.), which adapts difficulty based on responses. The reading section tested overall comprehension, grammatical form, and implied meaning, while the listening section evaluated comprehension through dialogues and monologues. The maximum score for each section was 120.

6.5.6. Data Analysis

Reliability indices were calculated, and descriptive statistics were examined. Normality of data distributions was assessed using Shapiro-Wilk tests, with detailed descriptive statistics and normality analyses available in the online materials. To explore interrelationships among variables, bivariate correlations and exploratory factor analyses were conducted. A multilevel modeling approach was used to determine the predictive power of language aptitude. The significance level was set at .05. Reliability and normality checks as well as factor analyses were performed using SPSS, version 29.0 (IBM Corp., 2023). Correlation tests and multilevel modeling were conducted using R studio, version 2024.04.1+748 (Posit team, 2024). A complete list of R packages used in the study is available in Appendix 3.

6.6. Results

6.6.1. Preliminary Analyses

Three factor analyses were conducted to examine how various measures cluster together: (1) measures of complexity, accuracy, and fluency from the speaking proficiency task; (2) LLAMA subtests and the SRT task; and (3) scores from the self-paced reading, elicited imitation, and gap-fill tests. The results of these analyses are presented in the following sections.

First, an exploratory factor analysis of the speaking proficiency measures was conducted using principal component analysis with direct oblimin (oblique) rotation, based on the expectation that the underlying factors would be interrelated (Tabachnick & Fidell, 2014). The assumptions for the analysis were checked (KMO = 0.61, Bartlett's test of sphericity = .001). The analysis revealed a single component with an eigenvalue above 1 ($\lambda = 2.64$), which accounted for 53% of the variance. This component was labelled speaking proficiency. The factor loadings for each speaking proficiency measure are illustrated in Figure 31, with a detailed overview provided in Table B in Appendix 4.

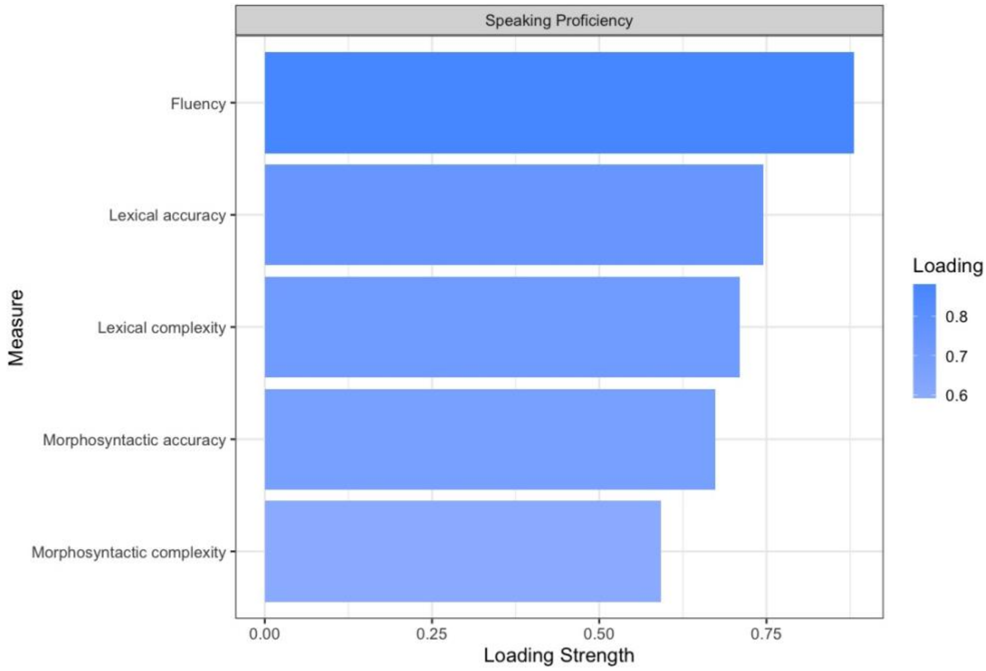

Figure 31. Factor loadings of speaking proficiency measures (for reference only; identical to Figure 21)

To identify the grouping of cognitive variables, a second exploratory factor analysis was performed using principal component analysis with direct oblimin (oblique) rotation. The assumptions were checked (KMO = 0.57, Bartlett's test of sphericity = .001). The analysis revealed two components with eigenvalues exceeding 1. LLAMA B, E, and F loaded on the first factor (λ = 1.66) which accounted for 33% of the variance and was labelled explicit learning aptitude (ELA). The SRT task and LLAMA D loaded on a separate factor (λ = 1.13) which explained 23% of the variance and was labelled implicit learning aptitude (ILA). Together, these two factors accounted for 56% of the total variance. The correlation coefficient between these latent factors was r = -0.097, which is well below the threshold of 0.8, indicating good discriminant validity (Brown, 2015). The factor loadings are depicted in Figure 32, with detailed information provided in Table A in Appendix 7.

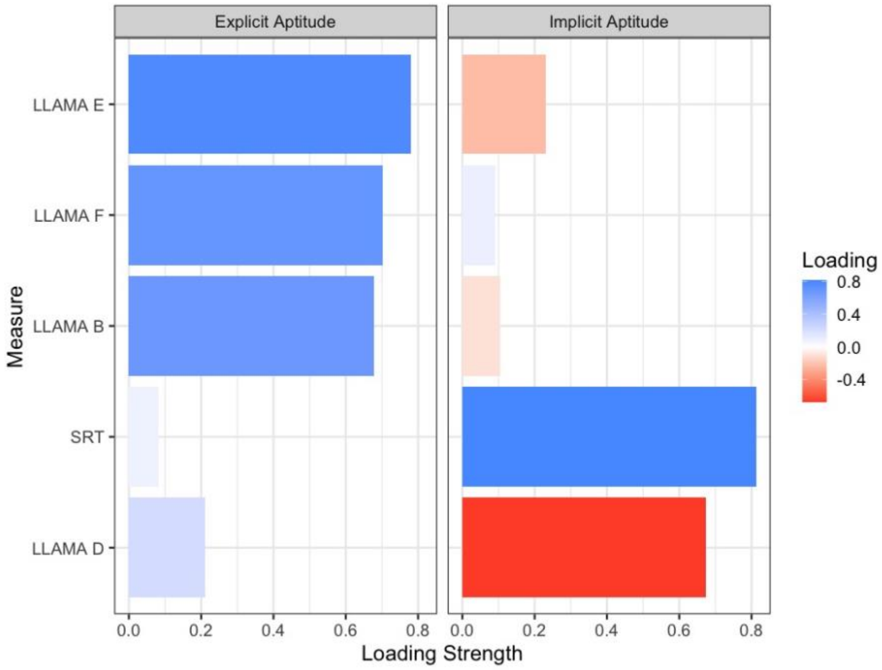

Figure 32. Factor loadings of the LLAMA subtests and the SRT task

6.6.2. How are scores on the self-paced reading, elicited imitation, and gap-fill tests related?

Bivariate correlations were run to examine the relationships between scores on the self-paced reading, elicited imitation, and gap-fill tests and establish how these measures relate to each other. Figure 33 shows the Spearman correlation coefficients in the upper triangle, density plots for each variable on the diagonal, and scatterplots for the variables in the lower triangle. The results indicate that scores on the self-paced reading task are weakly but significantly correlated with scores on the elicited imitation test (r_ρ = .269, p = .014). In contrast, a strong positive correlation is observed between scores on the elicited imitation and gap-fill tests (r_ρ = .860, p < .001). No significant relationship is found between self-paced reading scores and gap-fill scores.

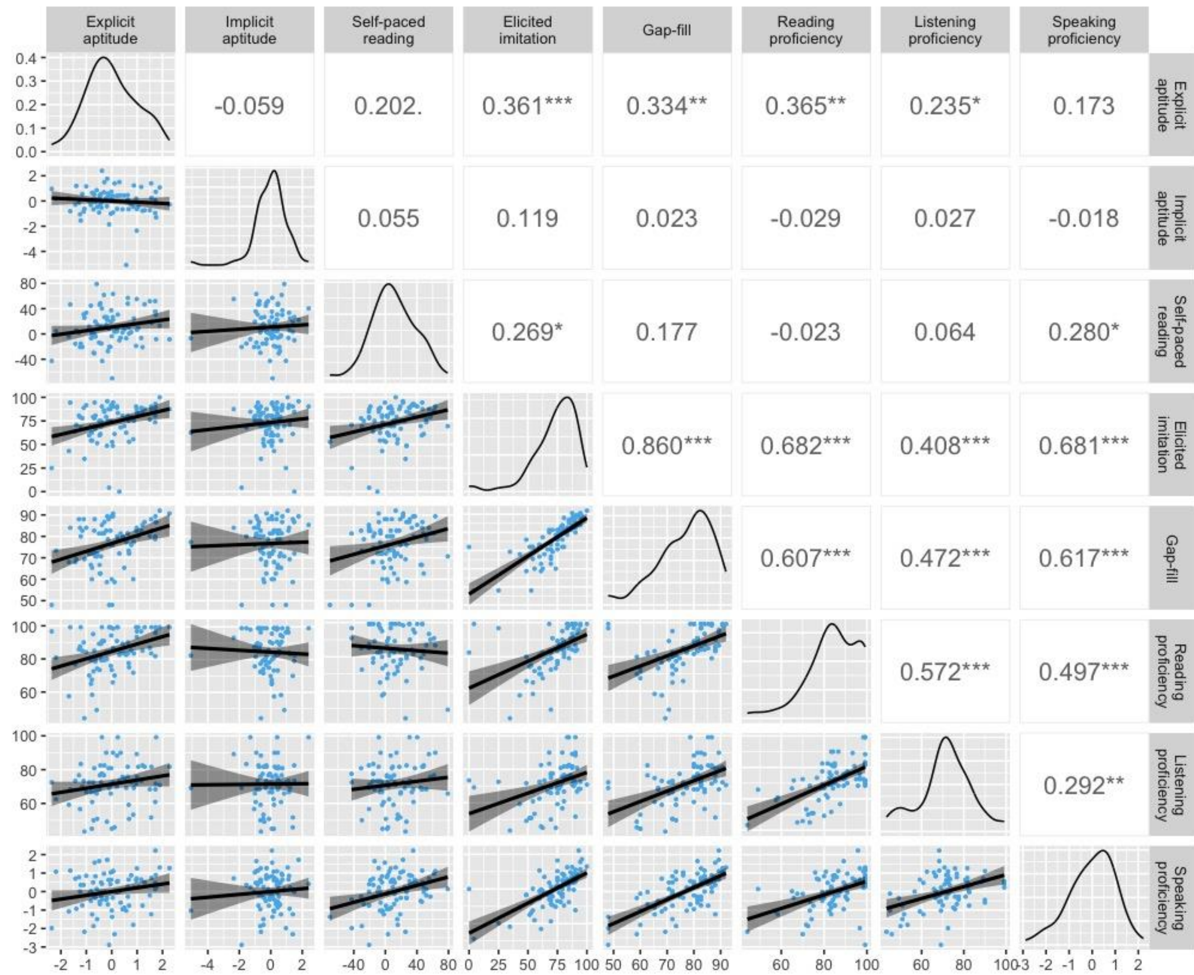

Figure 33. Correlations (Spearman's rho) between explicit and implicit aptitude, measures of L2 knowledge, and speaking, reading, and listening proficiency

To further explore how the three language variables cluster, an exploratory factor analysis was performed using principal component analysis with direct oblimin (oblique) rotation, based on the assumption that the underlying factors would be interrelated. The assumptions were checked (KMO = 0.53, Bartlett's test of sphericity = .001). Two components were identified with eigenvalues exceeding a .8 cut-off point as used in R. Ellis (2005). Elicited imitation and gap-fill tests loaded on a factor ($\lambda = 1.92$) that accounted for 64% of the variance, while the self-paced reading scores loaded on a separate factor ($\lambda = 0.90$) explaining 28% of the variance. Collectively, these factors accounted for an impressive 92% of the total variance, a substantial increase from the 75% reported by R. Ellis (2005). The correlation between the two latent factors was r = .296, well below the level of .80 (Brown, 2015) and .83 reported by Kim and Godfroid (2023), indicating good discriminant validity.

Figure 34 illustrates the factor loadings, with detailed information available in Table B in Appendix 7. The strong correlation between the gap-fill and elicited imitation scores, along with their high factor loadings on the first component, suggests that both measures tap into the same underlying construct of explicit knowledge. Conversely, the lack of correlation between the self-paced reading and gap-fill scores, combined with the loading of self-paced reading on a separate factor, indicates divergent validity. This supports the interpretation that the self-paced reading task measures a distinct construct of L2 knowledge, specifically implicit knowledge, as opposed to the explicit knowledge captured by the first factor.

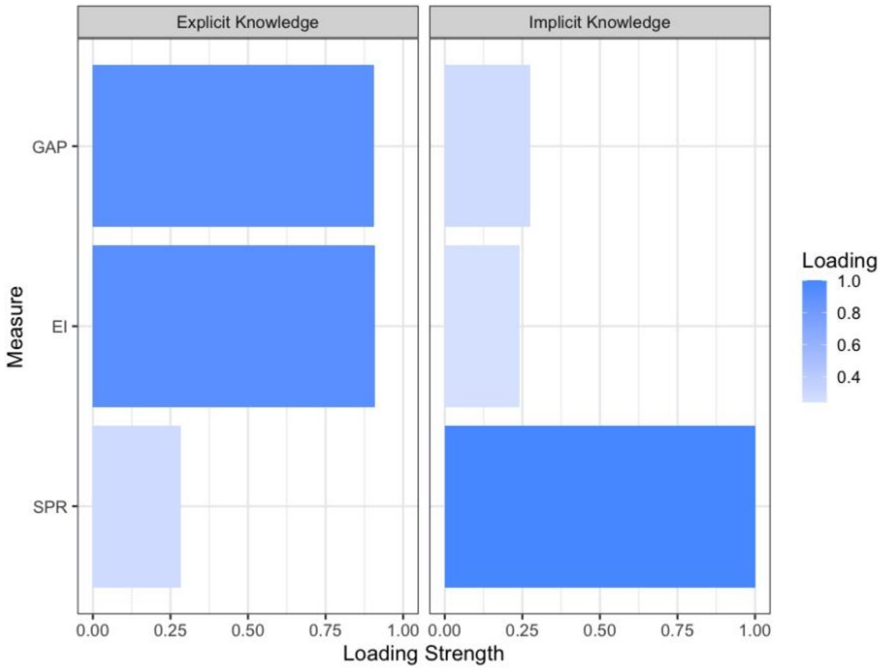

Figure 34. Factor loadings of measures of L2 knowledge

6.6.3. How do aptitude for explicit learning and aptitude for implicit learning contribute to learners' implicit knowledge?

To examine the relationships between explicit and implicit aptitude and implicit knowledge, a multilevel modeling approach was employed. This method allows for the inclusion of random effects across one or more grouping factors, which represent predictors that vary across different levels of a grouping factor within a regression model. This approach contrasts with fixed effects, which are assumed to remain constant across all levels of the grouping factor. The multilevel approach offers significant advantages over conventional linear regression, particularly in its ability to accommodate repeated-measures data. It is preferred over traditional linear regression because the inclusion of random intercepts and slopes enables models to effectively account for individual learner differences (for a detailed overview, see Cunnings & Finlayson, 2015).

A model was constructed with explicit and implicit aptitude as predictors, while reading, listening, and speaking proficiency were included as interacting variables (mediators), and implicit knowledge as measured by self-paced reading task was the outcome variable. All predictors were mean-centred. A detailed overview of the results is provided in Table A in Appendix 8, while Figure 35 offers a visual representation of the model.

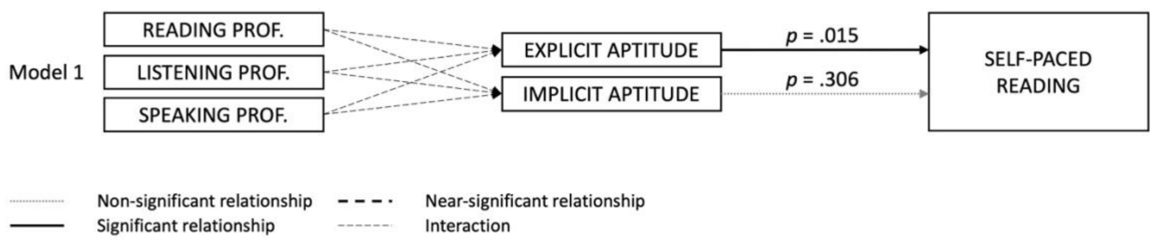

Figure 35. Multilevel model predicting self-paced reading scores

The results indicate that explicit aptitude significantly predicts implicit knowledge as measured by the self-paced reading task (model 1: estimate = 0.19, SE = 0.08, t = 2.460, p = .015). This finding provides direct evidence for a relationship between explicit aptitude and implicit knowledge. From this, we infer a relationship between explicit and implicit knowledge, suggesting an interface between the two.

6.6.4. How do aptitude for explicit learning and aptitude for implicit learning contribute to learners' explicit and automatized explicit knowledge?

To further explore the contributions of explicit and implicit aptitudes, two additional models were constructed, each with explicit and implicit aptitudes as predictors, and reading, listening, and speaking proficiency as interacting variables (mediators). In model 2, automatized explicit knowledge as measured by the elicited imitation test was the outcome variable. In model 3, explicit knowledge as measured by the gap-fill test was the outcome variable. A detailed overview of the results is available in Table B and Table C in Appendix 8, and Figure 36 is a visual presentation the models.

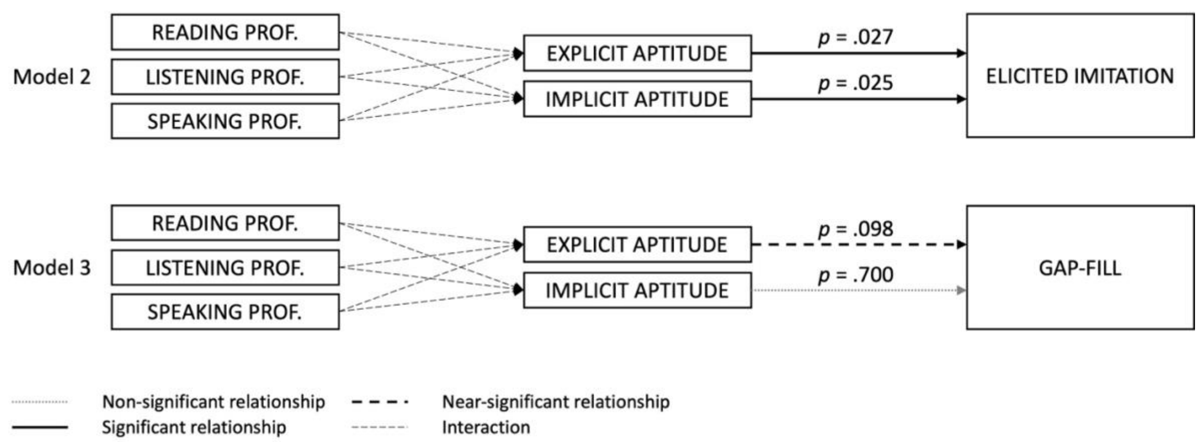

Figure 36. Multilevel models predicting elicited imitation and gap-fill scores

In model 2, explicit aptitude was found to significantly predict automatized explicit knowledge (estimate = 2.71, SE = 1.19, t = 2.68, p = .027). Additionally, implicit aptitude also significantly predicts automatized explicit knowledge (estimate = 3.24, SE = 1.39, t = 2.338, p = .025). The significance of both implicit and explicit aptitudes as predictors in this model suggests both an explicit-implicit interface and an implicit-explicit interface. In model 3, explicit aptitude emerged as a near-significant predictor of explicit knowledge (estimate = 1.81, SE = 1.08, t = 1.673, p = .098).

6.7. Discussion

The current study aimed to empirically investigate the relationship between aptitude for explicit and implicit learning and the corresponding knowledge types. To achieve this, bivariate correlations were run to establish interrelations, and a multilevel modeling approach was employed to explore the predictive relationships between aptitude and L2 knowledge.

In line with previous research, explicit language aptitude, which encompasses associative memory, phonetic coding ability, and language-analytic ability, influenced automatized explicit knowledge. The predictive influence of explicit aptitude on explicit knowledge is theoretically sound and empirically aligns with findings in Suzuki and DeKeyser (2017), and also Kim and Godfroid (2023) where the strongest predictor at any given time point was consistently the corresponding knowledge from an earlier time point. In light of Godfroid and Kim's (2021) assertion that a predictive relationship implies a shared cognitive process across tasks, the results of the current study are theoretically coherent in that explicit learning aptitude predicts the outcome of the learning process, i.e. explicit knowledge.

Correlation results revealed a significant relationship between measures of automatized explicit knowledge and implicit knowledge, thereby providing indirect evidence of an interface. More importantly, the results of a multilevel regression analysis showed a significant effect of explicit aptitude on implicit knowledge in model 1, providing direct support for an explicit-implicit interface. In addition, model 2 demonstrated a significant effect of implicit aptitude on automatized explicit knowledge, offering direct evidence of an implicit-explicit interface. Each of these findings will now be discussed in detail.

6.7.1. Explicit-implicit interface

The first key finding of the study emerged from model 1, which confirmed that explicit aptitude significantly predicted implicit knowledge, as measured by a self-paced reading task. In this task, implicit knowledge was calculated as a difference in reaction times between grammatical and ungrammatical items, with higher scores reflecting greater hesitation when encountering errors. This increased hesitation was positively associated with explicit aptitude, suggesting that individuals with stronger associative memory, language-analytic ability, and sound-symbol learning skills demonstrated heightened sensitivity to grammatical violations. This finding supports a predictive relationship, or interface, between explicit aptitude and implicit knowledge. It is in line with N. Ellis's (2005, 2006; 2015) interface position, which posits that the conscious registration of linguistic patterns in input, associated with explicit knowledge, can aid the implicit learning process by enabling further tallying of these patterns by implicit learning mechanisms.

The result corroborates findings from Kim and Godfroid (2023), who identified evidence of the explicit-implicit interface using a longitudinal design. Participants in Kim and Godfroid (2023) were immersed in a naturalistic setting as students living in a target-language country, with an average length of residence of 35 months at time point 1 and 36 months at time point 2. In contrast, participants in the current study averaged only 8 days in target-language countries, making for a substantial difference. The current study thus contributes to the interface debate by providing evidence in support of the explicit-implicit interface in a classroom setting with limited target language exposure.

The result also aligns with findings from a previous study by Suzuki and DeKeyser (2017) which sought to establish a relationship between explicit aptitude and implicit knowledge cross-sectionally in a naturalistic setting. Specifically, Suzuki and DeKeyser (2017) proposed that the interface is evidenced through an "indirect" facilitative role of explicit aptitude, where explicit aptitude influences automatized explicit knowledge, which subsequently impacts implicit knowledge. In contrast, the current study provides direct evidence of the explicit-implicit interface through the significant predictive relationship between explicit aptitude and implicit knowledge. Thus, this study not only corroborates findings from Suzuki and DeKeyser (2017) but further strengthens the evidence for the explicit-implicit interface.

An explanation for the discrepancy between findings of this study and Suzuki and DeKeyser's findings may lie in the mediating role of proficiency, which was accounted for in the current research but not in Suzuki and DeKeyser (2017), where only highly proficient L2 learners were included. Pavlekovic and Roehr-Brackin (under review) [Chapter 5] suggest that explicit aptitude is primarily relevant at earlier stages of L2 proficiency, echoing similar findings by Artieda and Muñoz (2016). Drawing on these findings, it is possible that explicit aptitude impacts implicit knowledge primarily until learners reach a certain level of L2 proficiency. Therefore, it is plausible that the participants in Suzuki and DeKeyser's study were too proficient for the direct effects of explicit aptitude to be observable.

In sum, the findings suggest that the explicit-implicit interface is a dynamic phenomenon, robust to different learning contexts, but with the strength of the association potentially depending on internal factors such as L2 proficiency.

6.7.2. Implicit-explicit interface

The second key finding of this study was that implicit aptitude impacted automatized explicit knowledge in model 2. Implicit aptitude, operationalized as a latent factor comprising implicit sequence learning ability and auditory pattern recognition ability, showed a significant predictive influence on elicited imitation performance. This type of relationship has been observed in only one other study (Kim & Godfroid, 2023) and never before in the context of aptitude-knowledge research using a cross-sectional design.

Implicit learning abilities appear to facilitate the development of automatized explicit knowledge, which aligns with Bialystok's (1994a, 1994b, 2001) position. Bialystok views language acquisition and use through the process of language analysis and control, whereby learners become more aware of certain structures as their knowledge becomes more analysed. In this perspective, knowledge evolves through the restructuring of mental representations, leading to a transition from implicit knowledge to a more analysed, or explicit, form. This line of argument opens up the possibility that awareness may emerge from knowledge that is initially acquired implicitly (Bialystok, 1994b, 2001). This claim is supported both by the findings arising from the present study and from other studies conducted in naturalistic (Kim & Godfroid, 2023) and laboratory settings (Andringa, 2020). For instance, Kim and Godfroid (2023) found that implicit knowledge at timepoint 1 influenced the development of explicit knowledge at timepoint 2, while Andringa (2020) reported the emergence of awareness in an implicit learning task.

Despite the direct influence of implicit aptitude on automatized explicit knowledge, implicit aptitude did not predict gap-fill test scores nor was implicit knowledge related to explicit knowledge. This suggests that the implicit-explicit interface is present only between implicit aptitude and explicit knowledge that has not yet reached a high level of analysis (i.e. automatized explicit knowledge), while highly analysed knowledge, as gauged by untimed tasks like gap-fill, does not interface with implicit aptitude or knowledge.

Cleeremans's radical plasticity thesis (2008, 2011) posits that language knowledge develops along a continuum, progressing from implicit (unanalysed) knowledge (stage 1) to explicit (analysed) knowledge (stage 2) and then automatized (less analysed) knowledge (stage 3). This implies that an interface could only exist between implicit and explicit knowledge, or stages 1 and 2 (Godfroid, 2023). However, findings from the current study suggest (1) a lack of direct relationship between implicit aptitude and explicit knowledge that would correspond to such an interface, and (2) a significant relationship between implicit aptitude and automatized explicit knowledge indicating an interface between stages 1 and 3. While speculative, these results nonetheless cast doubt on the viability of the radical plasticity model as a unifying framework for addressing the interface problem (Godfroid, 2023), as it does not readily account for the relationships detected in this study and findings in Suzuki and DeKeyser (2017).

6.7.3. Bidirectional interface

The traditional interface debate focused mainly on the explicit-implicit relationship with the goal of establishing the role of explicit knowledge in the development of implicit knowledge, while research in developmental psychology has often viewed language development through the lens of an implicit-explicit relationship. Results of the present study suggest that the two coexist – awareness facilitates implicit knowledge development, while it also emerges from implicit learning, creating a bidirectional interface between the two types of knowledge. Notably, the only other study to identify this bidirectional interface is Kim and Godfroid (2023), which involved university students averaging 27 years old at the start of the study. This contrasts with the younger, 16-year-old cohort in the present study, who are at the start of their cognitive maturity. Additionally, Kim and Godfroid's naturalistic setting provided participants with significantly greater target language exposure, as noted earlier. Despite these differences, both studies provide evidence supporting a bidirectional interface. The consistency of this finding across diverse settings underscores the robustness of the bidirectional interface and suggests the generalizability of this finding across the age ranges tested so far.

Lastly, when regression analyses were conducted separately for each grammatical structure, both explicit and implicit language aptitude remained significant predictors of elicited imitation scores, even for articles – a structure typically considered difficult to learn implicitly. In addition, explicit aptitude remained a significant predictor of self-paced reading scores for all three grammatical structures. This challenges the notion of structure-specific interfaces, as proposed by Han and Finneran (2014) and suggests that both the explicit-implicit and the implicit-explicit interface operate across a range of structures, regardless of their difficulty, though further research is required to confirm this across a broader set of grammatical features.

6.8. Conclusion

The findings of the present study highlight the coexistence of both an explicit-implicit and an implicit-explicit interface in language learning, suggesting a dynamic and bidirectional relationship between these two types of knowledge. On the one hand, the results support the traditional explicit-implicit interface, where conscious knowledge of linguistic patterns facilitates the development of implicit representations, aligning with the interface hypotheses proposed by scholars such as N. Ellis (2005, 2006; 2015), R. Ellis (1994; 2005, 2006), and Hulstijn (2015), who argue that explicit knowledge facilitates the transition to implicit knowledge, albeit with different explanations on how this is achieved. On the other hand, the study also provides evidence for the implicit-explicit interface – consistent with Bialystok's (1994a, 1994b, 2001) perspective from developmental psychology, where implicit knowledge evolves into explicit knowledge through developing mental representations grounded in the processes of analysis and control. Finally, the results indicate the robustness of the bidirectional interface across grammatical structures of varying difficulty, while comparisons with other studies point tentatively toward its stability across different age groups and learning contexts as well as variability due to the potential impact of proficiency.

6.9. Limitations and suggestions for future research

This study has several limitations that should be acknowledged, along with recommendations for future research. Firstly, the outcome measures were limited to three specific grammatical structures in a single L2, which restricts the generalizability of the findings and our ability to fully explore the impact of structure difficulty on the interface. Future research would benefit from incorporating a broader range of grammatical structures and other L2s to enhance the robustness of the findings. Secondly, while the study included a sufficient number of outcome measures to support the distinction between the constructs of explicit and implicit knowledge, a greater number of linguistic measures would have provided more statistical power to the analyses. Lastly, the sample size in this study was somewhat limited, which restricted statistical flexibility in terms of data analyses. Future studies, particularly those employing a cross-sectional design, should aim to recruit larger sample sizes to mitigate this issue.

7. Chapter 7 – Moderating role of structure difficulty

Research question nine aimed to investigate the role of structure difficulty in moderating the relationship between language aptitude and working memory on the one hand and gap-fill scores on the other hand. Addressing this research question required drawing on data and variables from the previously presented articles. Due to the complexity of the analyses and subsequent findings, it was not possible to incorporate this investigation into the existing articles, so it is presented in a separate chapter. The subsequent paragraphs provide a brief outline of the methodology as well as the results of a multilevel analysis used to analyse the role of structure difficulty.

The study incorporated a combination of cognitive and linguistic tasks. LLAMA B was used to measure associative memory, LLAMA E assessed phonetic coding ability, and LLAMA F measured language-analytic ability – each a component of explicit aptitude. In contrast, LLAMA D gauged auditory pattern recognition ability, and a serial reaction time task probed implicit sequence learning ability – both hypothesized to encompass implicit aptitude. Phonological working memory was assessed using a forward digit span task, while executive working memory was evaluated through an operation span task.

Linguistic measures included a self-paced reading task to assess implicit knowledge, an elicited imitation test to evaluate automatized explicit knowledge, and a gap-fill test to tap into explicit knowledge. Additionally, the elicited imitation task featured a word monitoring component designed to probe implicit knowledge. However, due to a substantial amount of missing data, scores from this component were excluded from the final dataset (see Section 3.4.2. for more details).

Difficulty was operationalized as objective difficulty, following the framework proposed by Housen and Simoens (2016) and DeKeyser (2003). Three target structures were selected to represent varying degrees of linguistic difficulty. Articles were considered the most difficult, due to the limited markedness of definiteness and high optionality of markers of definiteness in Croatian. The past tense was considered to be comparatively easier, owing to its higher perceptual salience, particularly in irregular verbs, lower communicative redundancy, and more transparent form-meaning mappings. The passive voice was viewed as more difficult than the past tense, due to its greater opacity, increased cognitive demands (especially in complex tenses), and higher communicative redundancy. However, it was deemed less difficult than articles, given the strong correspondence between the English and Croatian passive constructions. A detailed explanation with examples is provided in Section 3.2. Each outcome measure included items targeting all three structures. This design allowed for testing and comparing the predictive roles of implicit and explicit language aptitude and working memory across these structures, providing insight into any moderating effects of structure difficulty.

7.1. Results

7.1.1. Are the effects of explicit and implicit aptitude and working memory on learners' L2 knowledge of selected L2 morphosyntactic structures moderated by structure difficulty?

To address this research question, multilevel modelling was employed with participants treated as random variables, cognitive variables (explicit aptitude, implicit aptitude, working memory) as predictors, and self-paced reading task, elicited imitation and gap-fill test scores as outcome variables. Structure difficulty was included, allowing the examination of the main effects of explicit and implicit aptitude and working memory, as well as enabling the establishment of any moderating effects of structure difficulty. All predictors were mean-centred. Figure 37 provides a visual representation of the models, while a detailed overview of the results is provided in Appendix 9.

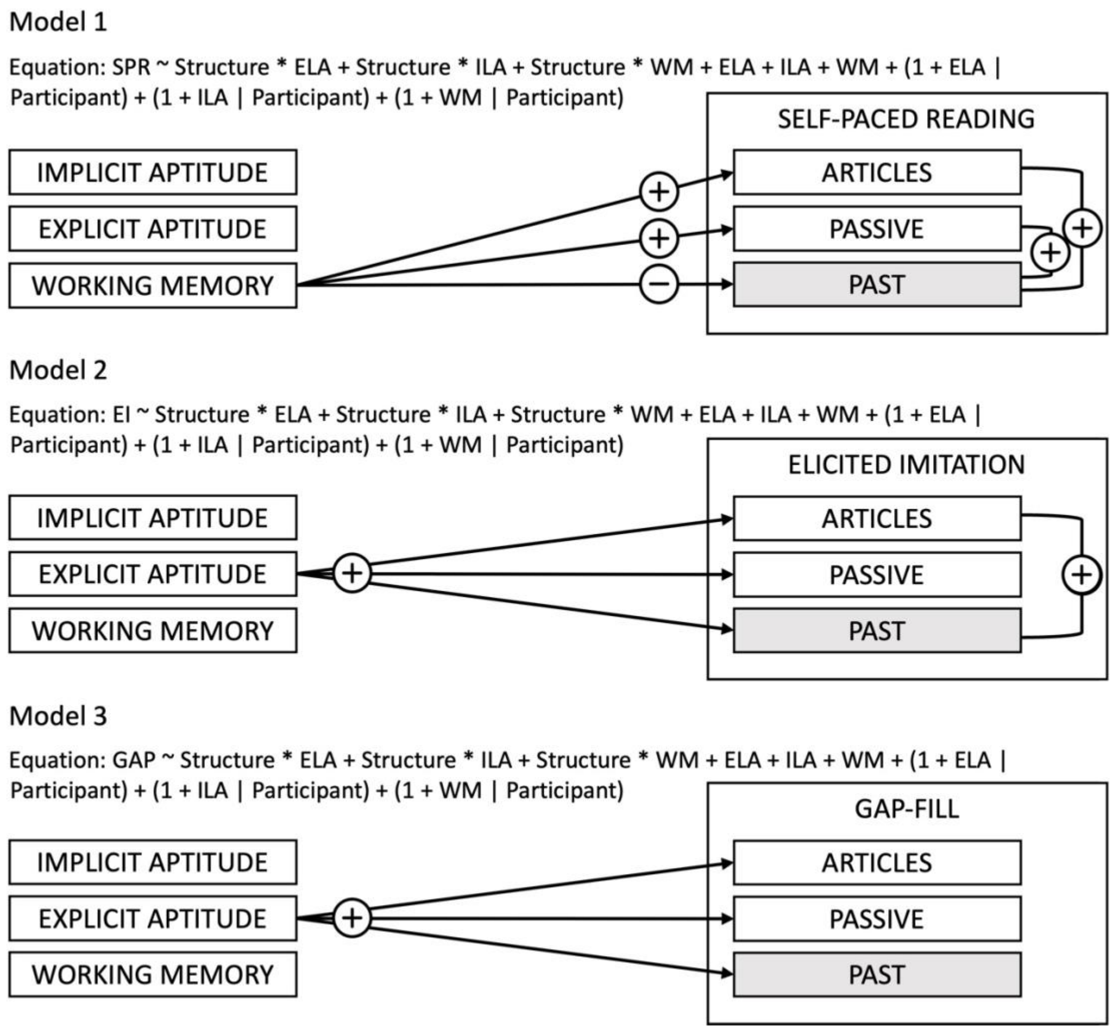

Figure 37. Multilevel models (Chapter 7)

7.1.1.1. Main effects

As shown in Figure 37, working memory had a significant negative main effect in model 1 (estimate = -0.24, SE = 0.11, t = -2.256, p < .0253). Additionally, explicit aptitude had a significant positive main effect in model 2 predicting elicited imitation (estimate = 5.22, SE = 1.49, t = 3.498, p < .0007) and in model 3 predicting gap-fill (estimate = 4.28, SE = 1.26, t = 3.388, p < .0009). These results indicate that while working memory negatively predicts implicit knowledge, explicit aptitude positively predicts explicit and automatized explicit knowledge. No other significant main effects were observed in the models.

7.1.1.2. Moderating effects of structure difficulty

Next, to explore whether the relationship between cognitive variables and L2 knowledge varied by structure difficulty, interaction terms between structure difficulty and each cognitive predictor (explicit aptitude, implicit aptitude, and working memory) were examined. A significant interaction would suggest that the relationship between cognitive abilities and explicit knowledge is influenced by structure difficulty. The analysis used the simple past structure as the reference level, so the main effect reported in Figure 37 above reflects the relationship for the simple past structure, while interaction terms indicate whether this relationship is statistically different for the passive and articles. Unlike some other statistical tools which compare each group to the grand mean of all groups to calculate the significance of differences, R directly compares group means. While the choice of reference level does not impact the overall results, it determines how the results are presented.

Model 1: Predicting implicit knowledge

There was a significant interaction between working memory and structure difficulty in model 1, which predicted implicit knowledge. Specifically, the negative relationship observed for the simple past (estimate = -0.24, SE = 0.11, $t = -2.256$, $p < .0253$) was significantly different from the relationships observed for the passive voice (interaction: estimate = 0.29, SE = 0.15, $t = 2.022$, $p < .0446$) and articles (interaction: estimate = 0.29, SE = 0.15, $t = 2.023$, $p < .0444$). The positive interaction estimates for both the passive voice and articles indicate that, unlike the negative relationship found for the simple past, working memory is a positive predictor for these two more challenging structures. This suggests that structure difficulty moderates the relationship between working memory and implicit knowledge.

In other words, working memory is a significant negative predictor for the easy structure (simple past), but becomes a positive predictor for the more difficult structures (passive and articles). To further investigate this relationship, two additional models were run with passive voice and articles set as the reference levels, respectively. In both cases, working memory was not a significant predictor (passive: estimate = 0.05, SE = 0.11, t = 0.469, p = .6397; articles: estimate = 0.05, SE = 0.11, t = 0.471, p = .6381), indicating that although the relationship changes direction depending on the target structure, it is not strong enough to be considered significant for these two (more difficult) structures. This pattern is illustrated in Figure 38, where the three relationships (simple past, passive, and articles) are shown in different colours for comparison.

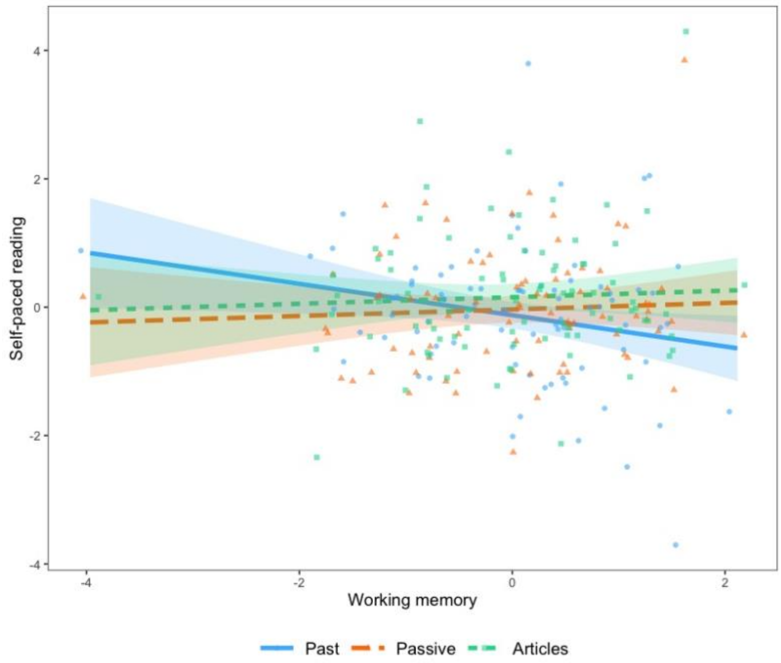

Figure 38. Relationship between implicit knowledge and working memory divided by linguistic structure

In conclusion, working memory is a significant negative predictor for the simple past, but it is not a significant predictor for the passive voice and articles, confirming the moderating role of structure difficulty in the relationship between working memory and implicit knowledge.

Model 2: Predicting automatized explicit knowledge

A significant interaction was observed between explicit aptitude and structure (interaction estimate for articles = 2.59, SE = 1.27, $t = 2.032$, $p < .0437$). This suggests that the positive relationship detected for the simple past (estimate = 5.22, SE = 1.49, $t = 3.498$, $p < .0007$) is significantly stronger for the more difficult structure, articles. To confirm, the model was rerun with articles as the reference level and explicit aptitude remained a significant positive predictor, but had a higher estimate (estimate = 7.81, SE = 1.49, $t = 5.230$, $p < .0001$). This indicates that as the structure difficulty increases, the strength of the predictive relationship between explicit aptitude and automatized explicit knowledge also increases. In contrast, no significant interaction was found between explicit aptitude and the passive structure in either model, suggesting that the relationship between explicit aptitude and automatized explicit knowledge for the passive structure does not differ significantly from that of the simple past or articles. The varying slopes of the predictive relationship between explicit aptitude and EI scores for the three structures can be seen in Figure 39, where the steepest slope signifying the strongest relationship is observed for articles, followed by the passive and the simple past, reflecting the decreasing strength in the relationship as difficulty decreases.

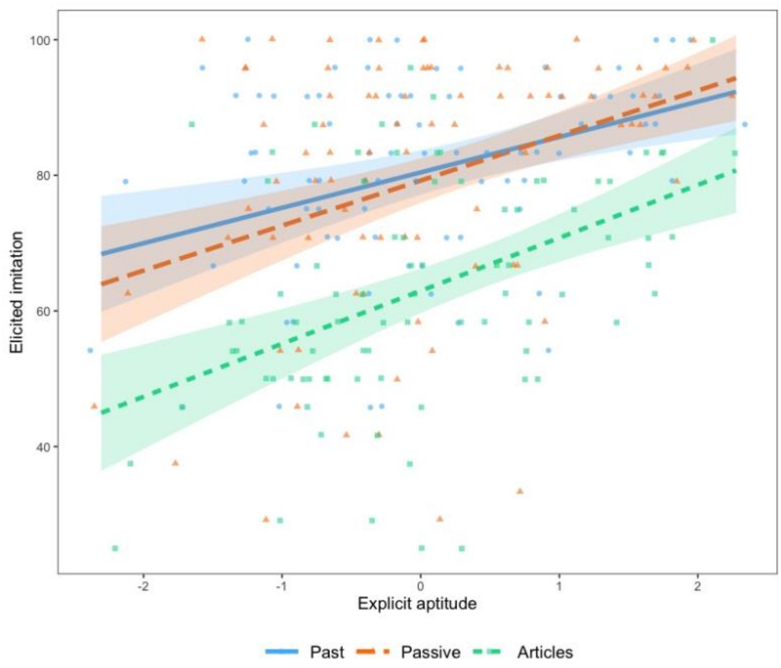

Figure 39. Relationship between automatized explicit knowledge and explicit aptitude divided by linguistic structure

Model 3: Predicting explicit knowledge

As noted, explicit aptitude was a significant positive predictor of gap-fill scores (estimate = 4.28, SE = 1.26, t = 3.388, p < .0009). However, no significant interaction terms between structure and explicit aptitude were found, indicating that the predictive relationship between explicit aptitude and explicit knowledge as measured by the gap-fill test is stable across the three structures. This suggests that the positive relationship between explicit aptitude and explicit knowledge, as operationalized by the gap-fill test, is not influenced by the difficulty of the structure being tested.

7.2. Discussion

The aim of the current chapter was to establish whether structure difficulty moderates the effects of explicit and implicit aptitude, as well as working memory, on L2 knowledge and how this relationship differs across different types of knowledge.

To address this question, the role of structure difficulty was examined using multilevel modeling, which produced three models – one for each of the outcome measures. The results pointed to two significant interactions: one between structure difficulty and the effects of working memory, and another between structure difficulty and explicit aptitude. In the model predicting implicit knowledge via the self-paced reading task, working memory was found to be a significant negative predictor for the simple past tense, while this relationship was neutralized for the passive voice and articles. This suggests that for the simplest morphosyntactic structure, higher levels of working memory capacity were associated with lower scores on self-paced reading.

Self-paced reading performance was operationalized using the grammatical sensitivity index, where higher scores indicate longer processing times for ungrammatical sentences, assumed to reflect greater grammatical sensitivity and representing implicit knowledge. Surprisingly, model 1 showed that participants with lower working memory experienced greater processing delays, while those with higher working memory processed sentences more quickly. This effect aligns with the Constrained Capacity Model (Just & Carpenter, 1992), which posits that individuals with higher working memory process ambiguous sentences more rapidly.

Since the GSI reflects the time difference required to process ungrammatical sentences compared to grammatical ones, it is plausible that ungrammaticality can result in ambiguity, particularly for learners who possess the implicit knowledge necessary to detect grammatical violations. Although ambiguity was not deliberately introduced in the stimuli, it seems reasonable to expect that higher working memory capacity could facilitate faster sentence processing and thus reading, especially given that participants were instructed to read for comprehension (Juffs & Harrington, 2011).

This casts doubt on the validity of the self-paced reading task as a pure measure of implicit knowledge, as the delays in reading ungrammatical sentences may simply result from processing constraints caused by ambiguity. In other words, the results suggest that the self-paced reading task, because it assesses knowledge at the processing level, may not isolate implicit knowledge without being influenced by an individual's working memory capacity. This explanation helps to clarify why accuracy-based measures sometimes demonstrate better predictive validity compared to reaction-time-based measures (Godfroid & Kim, 2021). It also highlights the growing need for additional validation studies of reaction-time measures that account for the role of working memory.

In examining passive and article structures, however, no relationship was found between GSI or processing speed and working memory. This outcome is desirable, as hesitation (or slower processing time) in tasks designed to measure implicit knowledge should ideally reflect sensitivity to grammatical violations rather than working memory capacity. This further informs the claims about self-paced reading as a measure of implicit knowledge: it appears that for simpler grammatical structures, working memory may obfuscate implicit knowledge, whereas for more complex structures, this effect does not seem to occur. Although a thorough syntactic processing analysis is beyond the scope of this thesis, these results underscore the need for further research to assess the validity of self-paced reading as an implicit knowledge measure for L2 learners and to explore how target difficulty or syntactic complexity might influence this relationship.

For the model predicting automatized explicit knowledge via the elicited imitation test, explicit aptitude was a significant predictor across all structures, with the strength of this relationship increasing with structure difficulty. This suggests that learners rely more on explicit aptitude when processing more challenging structures, supporting previous findings that learners engage their language-analytic abilities more with difficult structures (Robinson, 2002; Yalçın & Spada, 2016) and affirming the moderating role of structure difficulty.

Conversely, for the gap-fill test, structure difficulty did not moderate explicit aptitude effects; explicit aptitude remained a consistent positive predictor across all structures. A possible explanation for the different results between the elicited imitation and gap-fill tests lies in the nature of the knowledge each test assesses. The gap-fill test primarily taps into explicit knowledge, allowing participants to engage the task without time pressure and rely on rule-based knowledge with high confidence in their answers. In contrast, the elicited imitation test is believed to assess "a body of conscious linguistic knowledge including different levels of automatization" (Suzuki, 2017, p. 1230). It is therefore plausible to assume that participants' linguistic knowledge of easier structures was more automatized, reducing their reliance on explicit aptitude. Conversely, their knowledge of more difficult structures may have been less automatized, rendering greater need for reliance on explicit aptitude – an effect also observed in the gap-fill results.

Together, these findings present a complex picture. Working memory impacts performance on implicit knowledge tasks where its influence is less desirable due to the risk of confounding processing times used to calculate the grammatical sensitivity index. Explicit aptitude, by contrast, reliably predicts performance on tasks assessing both automatized and non-automatized explicit knowledge, irrespective of structural difficulty, although its influence intensifies for more difficult structures, especially in tasks requiring automatized explicit knowledge. This suggests that task characteristics significantly influence the role of language aptitude, in alignment with Robinson's dynamic view of aptitude (2005, 2007, 2012).

8. Chapter 8 – Relationship between accuracy, certainty, and source attributions

The final research question aimed to examine the relationship between source attributions, confidence ratings, and accuracy scores on a gap-fill test probing the knowledge of articles, passive, and past simple. The subsequent paragraphs provide a brief overview of the methodology as well as some theoretical and empirical considerations related to the nature of L2 knowledge probed by a gap-fill test.

The study employed a gap-fill which was a multiple-choice test that additionally incorporated subjective measures of awareness – confidence ratings and source attributions, i.e. whether learners relied on their knowledge of rules, memory, intuition, or whether they were guessing (Dienes, 2008; Dienes & Scott, 2005; Dienes et al., 2012). Each gap-fill item included a certainty scale where participants selected their confidence level on a scale from 50% to 100%. Participants were also asked to select a source attribution for each item, with the following options: rule, intuition, memory, and guess (see Section 3.3.3). By assessing both explicit knowledge (accuracy), and subjective measures of awareness (source attributions and confidence ratings), we can examine how participants approached the test and how their confidence and perceived knowledge related to their actual knowledge as measured by their performance on the test.

Reber (1967) and Reber et al. (1991) argued, on the basis of evolutionary theory, that implicit knowledge shows lower variability (see Section 2.1.1.1) compared to explicit knowledge, while (Tarone, 1988) claimed that learners' implicit knowledge is highly systematic. Subsequently, R. Ellis (2005) designed a framework for determining whether a task is gauging implicit or explicit knowledge. He outlined seven criteria by which the type of knowledge probed by a test can be determined. Among other things, he posited that implicit grammar knowledge is largely systematic and learners therefore should display systematicity on a test of implicit knowledge with little intra-individual variation. Explicit knowledge, on the other hand, is said to be more imprecise, inaccurate, and inconsistent (Sorace, 1985) with learners often having vague rather than clear knowledge of rules. Consequently, learners should display greater systematicity when deploying implicit compared to explicit knowledge, while they should also display greater certainty or confidence in their answers when they draw on implicit compared to explicit knowledge.

In an empirical investigation, R. Ellis (2005) asked L1 Chinese learners of L2 English to indicate the degree of certainty and whether they relied on “rule” or “feel” on a battery of 5 tests. Contrary to expectations, tests probing explicit knowledge such as untimed grammaticality judgment test and metalinguistic knowledge test were strongly and positively related to greater confidence. Confidence ratings on tests designated as tests of implicit knowledge were not related to confidence ratings. These findings cast doubt on the initial hypotheses about certainty, instead indicating that learners experience higher confidence levels on tests of explicit knowledge. Furthermore, greater reliance on rule knowledge was related to higher accuracy scores on tests hypothesized to assess explicit knowledge, but not on tests hypothesized to gauge implicit knowledge.

Moreover, Tomak (2019) utilized a gap-fill test to probe explicit knowledge of articles in L1 Russian learners of L2 English and utilized confidence ratings to examine the relationship between subjective measures and test performance. The study further expanded research into source attributions by introducing four possible sources: guess, intuition, rule, and memory. The results pointed to the positive relationship between accuracy and confidence, as well as between confidence and reliance on knowledge of rules thus indicating that learners are aware of their knowledge on a gap-fill test, their confidence is higher when they rely on rules, while it is lower when they guess and rely on intuition.

The current study will examine whether relationships between subjective measures of awareness and explicit knowledge assessed by a gap-fill test. The grammatical structures included are articles, passive voice, and past simple tense. In this context, the study aims to determine whether any counter-evidence to R. Ellis's (2005) framework found in Tomak (2019) applies to structures beyond articles.

8.1. Results

8.1.1. What is the relationship between source attributions, confidence ratings, and accuracy on a gap-fill test?

In order to investigate the interrelationships between source attributions, confidence ratings, and accuracy on the gap-fill test, descriptive statistics were first calculated by target structure. The results are available in Table 18.

Table 18. Source attributions and certainty ratings on the gap-fill test

	Articles		Passive		Past	
	M	SD	M	SD	M	SD
Rule	.470	.087	.481	.126	.474	.110
Memory	.212	.042	.187	.049	.195	.060
Intuition	.269	.061	.284	.079	.294	.058
Guess	.049	.033	.048	.053	.036	.025
Confidence	.904	.031	.910	.038	.903	.031
Accuracy	.769	.210	.887	.100	.795	.156

As can be seen, in the case of articles, passive, and past, participants showed similar behaviour and indicated that they relied most heavily on their knowledge of rules, followed by intuition, memory, and guess. For all three structures, participants reported high levels of confidence (>90%), while accuracy was likewise high with average scores across structures ranging from 77% to 89%.

In order to address the research question, bivariate correlations were calculated between accuracy, source attributions, and confidence ratings for each structure separately. This can be seen in Figure 40 below.

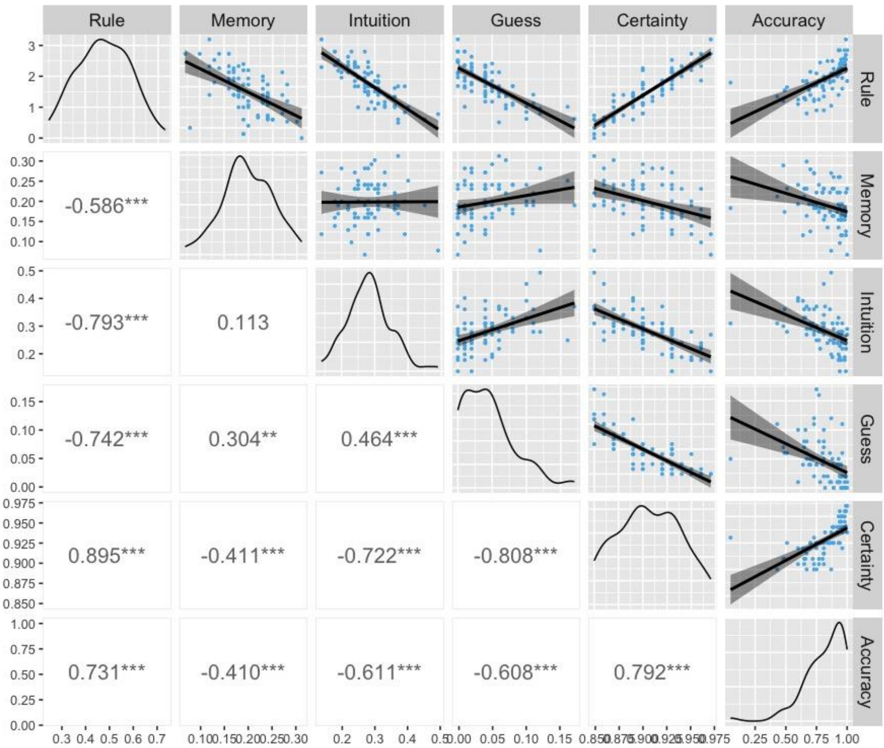

Figure 40. Correlations (Spearman's rho) between source attributions, certainty ratings, and accuracy of gap-fill scores

As illustrated in Figure 40, rule use was strongly and positively related to confidence ratings (r_ρ = .895, p = .001), indicating that participants were more confident when they indicated that they relied on rule as a source attribution. Memory, intuition, and guess were all significantly and negatively related to confidence ratings meaning that participants were significantly less confident when they selected one of the three source attributions. The strongest negative relationship was detected for guess (r_ρ = -.808, p = .001), followed by intuition (r_ρ = -.722, p = .001), and finally memory (r_ρ = -.411, p = .001) indicating that the largest negative effect on certainty was recorded when participants were guessing.

Accuracy was strongly and positively related to reliance on rule (r_ρ = .731, p = .001), while it was moderately strongly and negatively related to memory, intuition, and guess. This means that the more participants reported they relied on rule, the higher their accuracy was, while in the case of memory, intuition, and guess, the more they reported they relied on these, the less accurate they were. The strength of the negative relationship was similar for intuition (r_ρ = -.611, p = .001) and guess (r_ρ = -.608, p = .001), while reliance on memory generated somewhat less strong negative relationship (r_ρ = -.410, p = .001). Thus, accuracy and confidence were related to source attributions in a similar manner: reliance on knowledge of rules enhanced both, while reliance on memory, intuition, or guessing reduced confidence and accuracy. Finally, accuracy and certainty were strongly and positively related (r_ρ = .792, p = .001) indicating that participants were more accurate when they were more confident thus showing awareness of their knowledge. In other words, participants can judge their own knowledge well.

Next, partial correlations were calculated between source attributions and confidence ratings while controlling for accuracy to see if the previously reported relationships between source attributions and confidence are artefacts of their relationship with accuracy. All correlations remained in place – the relationship between rule use and confidence gained in strength (r_ρ = .839, p = .0001); while the relationships with between memory (r_ρ = -.260, p = .025), intuition (r_ρ = -.593, p = .001), and guessing (r_ρ = -.748, p = .001) and confidence lost in strength, but still remained significant. This means that reported use of rules is primary determinant of confidence ratings, above and beyond accuracy, with higher reliance on explicit knowledge of rules, related to higher confidence. In other words, participants were not more confident because their answers were accurate, but rather because they relied on their knowledge of rules. In the case of memory, the more participants related on their memory, the lower their confidence was, although the negative relationship was very weak. Both rule knowledge and memory are associated with explicit knowledge. However, in the case of memory, heavier reliance meant lower confidence.

R. Ellis (2005) postulated, on theoretical grounds, that explicit knowledge invokes low degree of certainty while implicit knowledge invokes high degree of certainty in responses. The results reported above, however, contradict this notion. There is a positive and significant relationship between the reported use of rules on the one hand, and accuracy on the other hand, meaning that learners are more accurate when they rely on their conscious knowledge of rules (i.e. explicit knowledge), while relying on intuition or guessing, source attributions more aligned with implicit than explicit knowledge, were significantly but negatively related to confidence meaning that learners' confidence was actually lower when they relied on these source attributions.

Further, bivariate correlation analysis showed the use of rule was associated with higher confidence indicating that participants were more confident when they relied on their knowledge of rules. Partial correlation analysis that controlled for the potential confounding effects of accuracy in this equation reaffirmed the findings – participants were significantly more confident when they reported rule use indicating that the main contributor to their confidence was indeed their reliance on the rules, above and beyond pure accuracy of the answer. Interestingly, in the case of memory, even after controlling for accuracy, the more participants relied on their memory, the lower their confidence was, although the negative relationship was very weak. This shows that rule knowledge and memory have different status in learners' minds – rule knowledge seems to be correlated to higher confidence, while the use of memory seems to be correlated with lower confidence. Knowledge of pedagogical rules that define form, function, and use of L2 features is highly schematic and generalizable and should be applicable across language and in different instances, while memory is assumed to represent a more specific representation that is based on a specific item retained by L2 learners.

Intuition and guess, on the other hand, are source attributions usually associated with implicit knowledge, i.e. when participants select intuition or guess, they are assumed to have relied on language knowledge that is beyond their conscious introspection and that is not verbalizable. Given the significant and negative relationship between relying on intuition and confidence as well as guessing and confidence, it is safe to assume that participants were significantly less confident the more they relied on their intuition and guessing, even after the effects of accuracy were controlled for. This means that reported reliance on intuition or guessing on the gap-fill task was the key driver of low confidence ratings.

To examine this further, a separate partial correlation analysis was conducted for each grammatical structure in which confidence ratings were probed for any relationship with reported use of the four source attributions. In the case of the articles, the use of rule still generated high positive relationship, while the use of intuition and guessing still negatively related to the confidence ratings. However, the negative relationship between memory and confidence disappeared and was no longer significant. Looking at the use of rule and memory, the two source attributions usually associated with explicit knowledge, this indicates that rule and memory influence learners differently as source attributions. While the use of rules seems to boost participants confidence in their responses, when they rely on their memory, their confidence is unaffected. In the case of the passive voice, the results mirrored that of the articles. When participants reported relying on their knowledge of rules, their confidence was higher, while when they reported relying on their intuition or guessing, their confidence was significantly lower. Their confidence was not related to their reliance on memory.

In this sense, the current study replicated the findings of Tomak (2019) and demonstrated that the same phenomenon – rule knowledge being associated with higher confidence, intuition and guessing with lower confidence, and memory with no relationship to confidence – extends to grammatical features beyond articles. Looking at the rule knowledge and memory as two source attribution associated with explicit or conscious knowledge, the findings highlight the different status of rule knowledge and memory in learners' minds. While memory is assumed to represent more specific, item-based knowledge reliant on particular occurrences of L2 constructions, rule knowledge is highly schematic, enabling generalization and application across multiple instances. These results could suggest that learners perceive high schematicity as more useful for tackling more difficult structures, such as articles and the passive voice.

Lastly, in the case of the simple past tense, the correlation coefficients remained largely unchanged compared to cumulative results with only small deviations in strength. Specifically, when participants were confronted with sentences testing their knowledge of the past simple, relying on rule knowledge generated greater confidence, while relying on memory or simply guessing generated lower confidence ratings. If participants relied on their intuition, on the other hand, their confidence scores were not affected as there was no relationship between using intuition and confidence. The results suggest that reliance on rule knowledge, a mechanism characterized by high schematicity as discussed above, yielded high confidence, while reliance on memory of specific instances negatively impacted confidence.

Interestingly, reliance on intuition did not produce the negative confidence effects observed for more difficult structures. Instead, there was a subtle shift, as reliance on intuition no longer undermined confidence, as was the case with more difficult grammatical structures. Assuming that reliance on intuition reflects the use of implicit knowledge, this may indicate that participants' implicit knowledge of the simple past was more developed compared to their knowledge of articles and the passive voice, possibly due to its higher schematicity or overall lower difficulty. Conversely, for more difficult structures, reliance on intuition may have negatively impacted confidence because implicit knowledge was less developed. Thus, it is plausible to assume that reliance on intuition can detract from learners' confidence when their implicit knowledge is underdeveloped, as seen with more difficult structures, but has no effect when such knowledge is more robust, as observed with the simple past.

9. General discussion

9.1. Introduction

This project investigated the effects of cognitive individual differences, specifically, language aptitude and working memory, on L2 knowledge. Chapter 4 (Article 1) focused on the structural aspects of language aptitude and the predictive power of cognitive abilities that constitute language aptitude for L2 proficiency. Chapter 5 (Article 2) examined the mediating role of proficiency in the effects of implicit and explicit language aptitude and working memory, while Chapter 6 (Article 3) addressed the question of the interrelation between implicit and explicit knowledge by exploring the relationship between language aptitude and L2 knowledge. Chapter 7 focused on the moderating role of structure difficulty in the effects of implicit and explicit language aptitude and working memory. Finally, Chapter 8 explored the relationship between explicit knowledge, certainty ratings, and source attributions in the gap-fill test.

The following paragraphs will summarize the key findings of each chapter and then discuss their theoretical implications in relation to (1) working memory, (2) explicit aptitude, (3) implicit aptitude, (4) the interface between explicit and implicit knowledge, and (5) theoretical models of aptitude.

9.2. Summary of findings

9.2.1. Chapter 4 (Article 1)

Chapter 4 investigated the architecture of language aptitude through exploratory factor analysis, using individual cognitive abilities as variables. The study aimed to determine whether working memory should be considered a part of language aptitude, whether a distinction exists between explicit and implicit aptitude, and if so, whether implicit aptitude, like explicit aptitude, is multicomponential. Hierarchical regression analysis was used to examine the predictive relationship between components of language aptitude and working memory with reading and listening proficiency, as well as morphosyntactic knowledge.

The results indicated that executive working memory overlapped with associative memory or ability to learn new vocabulary. The study further supported a distinction between implicit and explicit aptitude, as well as the treatment of working memory as separate from language aptitude. Explicit aptitude was found to consist of associative memory, phonetic coding ability, and language-analytic ability, thereby confirming existing research. Implicit aptitude was also shown to be componential, with implicit sequence learning ability loading positively and auditory sequence learning ability loading negatively on the same factor. This provided a new perspective in which aptitude for implicit learning is viewed as a cognitive proclivity with loading scores that vary according to how learners approach the task. The predictive nature of language aptitude was also confirmed: components of explicit aptitude positively predicted reading proficiency, while components of implicit aptitude negatively predicted it in a hierarchical regression model after accounting for explicit aptitude, indicating this effect if learners relied solely on implicit aptitude. As expected, grammar knowledge was heavily dependent on language-analytic ability, whereas listening proficiency was not predicted by any cognitive ability and was thus hypothesized to have developed as a consequence of development of reading proficiency and morphosyntactic knowledge.

9.2.2. Chapter 5 (Article 2)

Chapter 5 examined the mediating role of proficiency in the effects of language aptitude and working memory on L2 knowledge. To achieve this, multilevel modeling was used to assess the interaction effects between proficiency on the one hand, and language aptitude and working memory on the other. The outcome variables were implicit, explicit, and automatized explicit knowledge, assessed via self-paced reading, elicited imitation, and gap-fill tests, respectively. Speaking proficiency was measured through an oral production task evaluating complexity, accuracy, and fluency. The study aimed to pinpoint the exact proficiency level at which the effects of implicit and explicit aptitude and working memory shift from significant to non-significant, or vice versa.

The results showed that working memory was a negative predictor of L2 knowledge, consistent with findings from Chapter 4. However, this negative effect decreased as proficiency increased. Explicit aptitude was a positive predictor of L2 knowledge, with proficiency serving as a mediator that weakened these effects; in other words, as proficiency increased, the positive influence of explicit aptitude diminished. A speaking proficiency level equivalent to M + 1SD was identified as the tipping point: for proficiency levels below this threshold, explicit learning aptitude was a positive predictor, while for levels above it, the positive effect of explicit aptitude was no longer significant. Implicit aptitude was also a positive predictor of L2 knowledge, with proficiency again acting as a mediator. In this case, higher levels of proficiency amplified the positive effect of implicit aptitude. Thus, the influence of implicit learning aptitude grew stronger as proficiency increased. The tipping point for implicit aptitude was around the mean proficiency level, indicating that the beneficial effects of implicit aptitude were not evident below the mean, but became increasingly important above it.

9.2.3. Chapter 6 (Article 3)

Chapter 6 explored the relationships between aptitude for implicit and explicit learning on one side, and implicit, explicit, and automatized explicit knowledge on the other. In doing so, it aimed to contribute to the interface debate by examining whether there is an interface between implicit and explicit learning and knowledge. Given the causal nature of the interface question, the analysis included measures predictive of language learning success – aptitude for implicit and explicit learning – and measures representing the outcomes of learning – implicit, explicit, and automatized explicit knowledge. Due to the mediating role of proficiency in the effects of language aptitude identified in Chapter 5, speaking, reading, and listening proficiency were included as covariates to control for any mediating effects.

The findings suggested the existence of both explicit-implicit and implicit-explicit relationships, revealing a bidirectional interface. The explicit-implicit interface indicated that explicit learning and knowledge can contribute to the development of implicit knowledge. Furthermore, this relationship suggested that explicit knowledge acquired in the context of explicit instruction, such as in an EFL classroom, can facilitate the development of implicit knowledge. The implicit-explicit interface suggested that explicit knowledge, can emerge from implicit knowledge, a phenomenon more commonly discussed in research on children's language learning. An analysis of structures with varying learning difficulty pointed to the robustness of this interface across different structures. Comparisons with similar studies highlighted the potential influence of proficiency, as well as the stability of the interface in relation to learning context and age differences.

9.2.4. Chapter 7

Chapter 7 examined the moderating role of structure difficulty in the facilitative effects of language aptitude and working memory. Using multilevel modeling, which accounts for individual learner differences, the chapter explored the predictive relationships between implicit and explicit aptitude as well as working memory, on implicit, explicit, and automatized explicit knowledge. Implicit knowledge was assessed via self-paced reading, automatized explicit knowledge through elicited imitation, and explicit knowledge using a gap-fill test.

The findings indicated that the negative effect of working memory was present only for the easier grammatical structure (i.e. simple past), and this effect did not extend to more difficult structures (i.e. passive voice, articles). The results also showed that the facilitative effect of explicit aptitude was moderated by the difficulty of the grammatical structure. While explicit aptitude emerged as a significant predictor across all structures, it was a stronger predictor for more difficult structures. However, this was the case only for automatized explicit knowledge; in contrast, the predictive relationship for explicit knowledge remained unaffected by the difficulty of the target structures. Similarly, implicit knowledge was not moderated by the difficulty of the linguistic feature.

9.2.5. Chapter 8

Chapter 8 examined the nature of relationship between accuracy, confidence, and source attributions as reported in the gap-fill test. Results from subjective measures, such as confidence ratings and source attributions, were compared to objective measures, specifically test accuracy, and their interrelationship analysed to determine whether reliance on rules, or explicit knowledge, rendered lower confidence ratings, as suggested by R. Ellis (2005), or higher confidence, as indicated by a more recent empirical study by Tomak (2019). The chapter also aimed to identify whether these relationships varied depending on the difficulty of the grammatical structure.

The findings indicated that accuracy was positively related to confidence suggesting that learners were aware of their knowledge and highlighting the explicit nature of the test. After controlling for accuracy effects, rule use was associated with higher confidence across structures, indicating that higher confidence was linked to reliance on rule-based knowledge beyond mere accuracy. Conversely, guessing was associated with lower confidence across all structures. Intuition was related to lower confidence for the two more difficult structures but not for the easier structure. Similarly, memory was negatively related to confidence for the easier structure; however, this predictive relationship was neutralized for the more challenging structures. These results suggest that higher schematicity was perceived as more important than memory of specific instances for more difficult structures. Meanwhile, for the easier structure, the negative effect of intuition on confidence was neutralized, potentially due to higher implicit knowledge associated with the relative lower difficulty of the structure.

9.3. Working memory

The importance of working memory (WM) in L2 processing, learning, and overall proficiency is well established (Juffs & Harrington, 2011; Linck et al., 2014). Specifically, WM has been shown to be critical for reading comprehension (Alptekin & Erçetin, 2010; Leeser, 2007; Tyler, 2001), L2 writing (Abu-Rabia, 2001; Adams & Guillot, 2008), vocabulary acquisition (Masoura & Gathercole, 2005; Papagno & Vallar, 1995; Service, 1992; Service & Kohonen, 1995; Speciale et al., 2004), and L2 grammar learning (Biedroń et al., 2022; Grey et al., 2015; Kempe & Brooks, 2008; Kempe et al., 2010; Pawlak & Biedroń, 2021; Sagarra, 2017). However, there is also evidence suggesting that the role of working memory could be influenced by factors such as L2 proficiency (Serafini & Sanz, 2016). The following paragraphs will discuss the explanatory power of WM in L2 learning, the mediating role of proficiency, as well as the status of working memory compared with aptitude.

9.3.1. The relationship of working memory and language aptitude

While numerous studies have highlighted the role of working memory in L2 outcomes, the relationship between working memory and language aptitude has received comparatively little attention in SLA (Li, 2015). The current study aimed to address this gap, with results indicating a relationship between working memory and the LLAMA B subtest of the LLAMA battery. Specifically, executive working memory, operationalized as the ability to memorize and recall a series of letters while simultaneously solving simple math problems, was shown to be associated with associative memory, or the ability to learn novel lexical items. This finding aligns with previous findings by Li (2016), who reported similar associations in a meta-analysis with a sample of 13,000 L2 learners. Such a relationship is unsurprising, as vocabulary acquisition typically involves a reliance on memory. The correlation strength ($r = 0.273$) indicates approximately a 7% overlap between the constructs.

In contrast, phonological working memory did not correlate with any component of language aptitude. While this result is consistent with some studies (Hummel, 2009; Li, 2016), other research found contrasting evidence (Juffs & Harrington, 2011). Given the shared variance of 18% ($r = 0.425$) between executive and phonological working memory, along with the observed association between novel lexical learning ability and executive, but not phonological working memory, it is reasonable to conclude that executive working memory is implicated in vocabulary learning, at least in the sample of the present study. Additionally, previous research suggests that phonological working memory may play a more significant role at lower proficiency levels (Hummel, 2009; Linck et al., 2013; Serafini & Sanz, 2016). Thus, it is possible that the sample in the current study surpassed the proficiency range where phonological working memory is relevant.

While the correlational results provide evidence of some overlap between working memory components and language aptitude, suggesting that working memory may be partially represented in explicit aptitude tests, the two working memory components measured – executive working memory and phonological working memory – loaded together on a separate factor from aptitude components. This finding aligns with prior research (Granena, 2013a; Hummel, 2009; Roehr & Gánem-Gutiérrez, 2009b; Yalçın & Spada, 2016), to the extent that comparisons can be made. Overall, these results suggest that, despite some overlap, working memory is a distinct construct from language aptitude. Consequently, studies adopting an explanatory or predictive approach to the role of working memory for L2 learning would benefit from separate measures of language aptitude and working memory.

9.3.2. The predictive relationship of working memory with L2 outcomes

First, executive working memory did not emerge as a predictor for either reading or listening proficiency. Second, phonological working memory showed a negative predictive relationship with explicit knowledge (referred to as "grammar knowledge" in Chapter 4) as measured by a gap-fill test. Although there is no obvious explanation for this result, statistically, language-analytic ability was the primary predictor of grammar knowledge in the same model and was accounted for first. This suggests that the negative predictive relationship between phonological working memory and grammar knowledge is present primarily when learners rely on working memory alone, or when working memory is used without the support of language-analytic ability.

Following the factor analytic results, working memory was operationalized as a construct encompassing both phonological and executive component. Subsequent analysis showed no relationship between working memory and either overall proficiency or knowledge of the three grammatical features assessed through self-paced reading, elicited imitation, and a gap-fill test. However, once proficiency was included as a covariate in the multilevel model, the combined construct of working memory became a significant predictor of explicit knowledge measured through a gap-fill test. Notably, this predictive relationship was evident only at lower proficiency levels. At these low levels of proficiency, working memory had a negative effect on explicit knowledge; as proficiency increased, the negative effect of working memory diminished until it was no longer significant. Thus, proficiency served as a mediator, gradually neutralizing the negative impact of working memory as proficiency improved. Notably, the current findings align with prior research indicating a mediating role of proficiency for working memory (Serafini & Sanz, 2016); however, the observed negative direction in the SLA literature is rare, having been documented only once before (Linck et al., 2013). It is possible, as suggested by Linck et al., that high levels of working memory might allow for increased switching capacity, leading to an overreliance on L1 during the gap-fill task, which could have negatively affected task accuracy.

When the moderating role of structure difficulty was examined, a negative predictive relationship emerged between working memory and implicit knowledge for the easier structure, while this relationship was neutralized for more difficult structures. Higher scores on the implicit knowledge test indicated greater hesitancy, typically interpreted in SLA as grammatical sensitivity; however, the results in Chapter 6 suggest that such hesitancy could also reflect increased processing time due to limited working memory capacity. These findings imply that higher working memory facilitates faster processing and thus shorter hesitancy, but only for simpler structures. For more difficult structures, this advantage disappears, indicating that working memory does not enhance processing in these cases.

As this line of argument is derived from a single study, it suggests the need for further inquiry to confirm the finding and to better understand how working memory may influence use of implicit knowledge. If replicated, these results could clarify how working memory affects task performance in implicit knowledge assessments, particularly in self-paced reading tasks. The current results imply that self-paced reading may serve as a valid measure of implicit knowledge specifically for more difficult structures, where lower working memory does not constrain processing, thereby avoiding a conflation of lower working memory with implicit knowledge. By contrast, structure difficulty did not moderate predictive relationship with explicit or automatized explicit knowledge measures, which may be unsurprising given that these accuracy-based measures are designed to assess knowledge during the production rather than the comprehension phase.

In sum, on the componential level, executive working memory appears to overlap partially with the ability to learn novel lexical items, whereas phonological working memory, contrary to expectations, does not seem to play a significant role. Based on previous research, this may be because the higher overall proficiency of the sample surpassed the threshold at which individual differences in phonological working memory typically matter. Moreover, while there is some overlap between executive working memory and explicit aptitude, both executive and phonological working memory converge to form a construct distinct from language aptitude.

When examining the predictive effects of the working memory construct, negative relationships with language outcomes were observed. For explicit or conscious knowledge, proficiency mediated these negative effects, with such effects evident only at lower proficiency levels and neutralized at higher levels. For implicit or unconscious knowledge, the negative effects of working memory were present only for easier grammatical structures and diminished for more difficult ones. This pattern reflects the influence of limited working memory capacity on the processing of linguistic structures, which is problematic given the assumption that greater hesitation in such tasks reflects implicit knowledge. If limited working memory can interfere with implicit knowledge on tasks such as self-paced reading and word monitoring only for certain structures, as suggested in the current study, it may help explain the conflicting results regarding whether reaction-time- or accuracy-based measures are best suited to capture implicit knowledge. These findings underscore the need for further empirical research into the validity of reaction-time-based measures of implicit knowledge in SLA.

9.4. Aptitude for explicit learning

Aptitude for explicit learning has consistently emerged as one of the most reliable predictors of language learning outcomes, showing a medium effect size on average, as evidenced by a meta-analysis spanning 50 years of research (Li, 2015). In addition, some researchers propose that the effects of explicit aptitude may ultimately depend on factors such as learner proficiency level (Artieda & Muñoz, 2016) and the difficulty of the target structure (Yalçın & Spada, 2016). The following paragraphs will discuss the componential nature of explicit aptitude, its predictive and explanatory power in language learning, and the factors that mediate and moderate its beneficial effects.

9.4.1. Effects of components of explicit aptitude

The current study reaffirmed the multicomponential nature of explicit learning aptitude. Results indicate that associative memory, phonetic coding ability, and language-analytic ability all contribute to explicit aptitude, aligning with theoretical frameworks (Carroll, 1990; Skehan, 1998) and empirical evidence (Granena, 2012, 2013a; Roehr-Brackin et al., 2023). Each component loaded positively on the explicit aptitude factor, indicating its contributing role in explicit aptitude overall. Importantly, explicit language aptitude was shown to be distinct from working memory, despite associative memory partially overlapping with working memory. It was further revealed that explicit aptitude should be differentiated from implicit aptitude.

The analysis of individual abilities demonstrated that language-analytic ability is crucial for developing explicit L2 grammar knowledge, which is unsurprising given that grammatical learning involves understanding word functions and relationships within sentences. A detailed examination of both subjective and objective measures on the gap-fill test revealed that learners were highly aware of the knowledge they applied during the task. Notably, high confidence was linked not only to accuracy but was also shown to depend on a reliance on rule-based knowledge.

Associative memory, which underlies the ability to acquire novel lexical items, and phonetic coding ability, which supports linking sounds to symbols, both predicted reading proficiency. Reading proficiency was measured by comprehension of short paragraphs that assessed grammatical form, meaning, implied meaning, and overall comprehension. This aligns with expectations, as grapheme-phoneme mappings and lexical knowledge are fundamental to reading skills.

9.4.2. Effects of explicit aptitude as a multicomponential construct

Following the results of factor analysis, associative memory, phonetic coding ability, and language-analytic ability were treated as a single factor that was able to predict both explicit and automatized explicit knowledge of the three targeted structures. This is in line with Li's meta-analytic findings (2015) according to which explicit aptitude shows medium effect on language outcomes. More interestingly, after adding speaking proficiency as a mediator in the three multilevel models, the beneficial effects of explicit aptitude were shown to exist only at lower proficiency levels. As proficiency increased, the facilitative effects of explicit aptitude decreased until the tipping point of M+1SD was reached after which explicit aptitude no longer predicted L2 knowledge. This was true in the case of both automatized explicit and explicit knowledge. Despite the fact that there was no significant predictive relationship between explicit aptitude and implicit knowledge, the mediating effect of proficiency was still detected at non-significant level indicating that explicit aptitude exerted a stronger influence at lower levels of proficiency.

These findings contributes to a limited number of previous empirical studies showcasing the same phenomenon (Artieda & Muñoz, 2016; Morgan-Short et al., 2014) and a comprehensive meta-analysis (Li, 2015) in which it was concluded that younger, but more importantly less proficient learners are more likely to draw on their explicit aptitude. This is also in line with a claim by Carroll (1990) that (explicit) aptitude is most relevant when learning a language *ab initio.*

The benefits of high explicit aptitude for explicit knowledge, as gauged on a gap-fill test, were unaffected by the difficulty of linguistic structures; in other word, structure difficulty did not moderate effects of explicit aptitude on explicit knowledge. In the case of automatized explicit knowledge, however, the facilitative effects of explicit aptitude were boosted by structure difficulty – more difficult structures prompted greater reliance on explicit aptitude. This result underscores findings from other studies examining the role of structure difficulty (Robinson, 2002; Yalçın & Spada, 2016) and has several implications. On one level, the mixed findings regarding the impact of structure difficulty on the facilitative role of explicit aptitude may be attributed to the variety of L2 outcome measures used; different tests measure different types of knowledge, potentially leading to varied results based on the knowledge type assessed. On another level, the results suggest that the role of explicit aptitude is influenced by the input and the type of knowledge being engaged. Difficult linguistic features on tasks probing automatized knowledge appear to activate reliance on explicit aptitude. In contrast, for less challenging features or tasks involving highly analysed explicit knowledge without time constraints, the influence of explicit aptitude diminishes.

These findings also have implications for language teaching. A recent study by Roehr-Brackin et al. (2024) showed that deductive instruction can have a levelling effect at the earliest stage of adult language learning by neutralizing aptitude differences, enabling learners with lower aptitude to perform comparably to those with higher aptitude. Corroborating results exist with earlier interaction studies too (Erlam, 2005; Hwu et al., 2014; Hwu & Sun, 2012; Li et al., 2019; Sanz et al., 2016). The results of the current study thus suggest that deductive teaching may only be necessary at lower proficiency levels before learners reach the tipping point. Furthermore, the results indicate that deductive teaching may be especially effective for complex linguistic structures, where learners rely more on explicit aptitude and in cases where highly automatized explicit knowledge is essential.

9.5. Aptitude for implicit learning

Aptitude for implicit learning represents a set of cognitive abilities that support implicit L2 processing, learning, and use without conscious awareness (Granena, 2020; Li & DeKeyser, 2021). Reber's (1967) pioneering work on implicit learning in adults challenged earlier assumptions about a narrow range of variation in implicit learning abilities among learners, thereby bringing implicit learning research into focus for cognitive psychologists. Kaufman et al. (2010) further demonstrated that this variation can predict L2 learning outcomes, solidifying the idea of separate aptitudes for implicit and explicit learning within SLA. The following paragraphs discuss the distinct nature of implicit versus explicit aptitude, outline the structural properties of implicit aptitude, and examine its explanatory and predictive power in language learning success. Lastly, the mediating role of L2 proficiency and moderating role of structure difficulty in the facilitative effects of implicit aptitude is addressed.

9.5.1. The structure of implicit language aptitude

The results of this thesis provide strong support for the existence of distinct aptitudes for implicit and explicit learning, aligning with recent theoretical (Granena, 2020; Li, 2022; Li & DeKeyser, 2021) and empirical (Granena, 2013a; Roehr-Brackin et al., 2023) accounts. In this study, implicit aptitude was conceptualized as comprising implicit sequence learning ability and auditory pattern recognition ability. Both abilities loaded onto the same factor, separate from cognitive abilities associated with explicit aptitude and working memory. However, implicit sequence learning ability and auditory pattern recognition ability showed no correlation, a finding that, while surprising, is consistent with prior studies (Buffington et al., 2021; Godfroid & Kim, 2021; Li & Qian, 2021). One possible explanation lies in the modality differences between the tests: while the SRT task is visual and non-verbal, the LLAMA D task is auditory and verbal. Research in cognitive psychology suggests that accuracy across verbal and visual stimuli may not be directly comparable, and neurocognitive studies indicate distinct processing pathways for these types of stimuli. Additionally, the SRT task is a process-oriented measure that assesses on-task learning, while LLAMA D (v.2), is accuracy-based and evaluates the product of learning (Christiansen, 2019).

Interestingly, the two components pulled in opposite directions in terms of their factor loadings; implicit sequence learning ability loaded positively, while auditory pattern recognition ability loaded negatively on the same factor. This finding, consistent with another study (Iizuka & DeKeyser, 2023), suggests that implicit aptitude may be fundamentally different in nature from explicit aptitude. Li and DeKeyser (2021) propose that implicit learning may occur through multiple pathways, leading to abilities that do not necessarily correlate or overlap. As discussed in Chapter 4, the lack of relationship and opposing loadings may result from participants approaching the tasks similarly in line with their individual proclivities, despite the tasks requiring somewhat different strategies. For instance, the SRT task requires participants to engage with visual stimuli, with task success depending on their ability to process input without focal attention, as the pattern follows a second-order conditional sequence. In contrast, the LLAMA D task requires focal attention to memorize a series of auditory stimuli for later recognition. Here, focal attention is crucial, and attempting to apply the same cognitive strategies across tasks may enhance performance in one but hinder it in the other. This divergence may explain why the abilities do not correlate or load in the same direction.

These findings suggest that the traditional view of implicit aptitude, where higher levels are invariably seen as advantageous, may need reconsideration. SLA researchers may benefit from viewing implicit aptitude as a cognitive proclivity that varies across individuals.

9.5.2. Implicit aptitude as a predictor of L2 success

Interestingly, the two implicit aptitude components negatively predicted reading proficiency in a hierarchical regression model, in contrast to two components of explicit aptitude, associative memory and phonetic coding ability, which positively contributed to reading proficiency. Together, implicit aptitude accounted for 12% of the total variance in reading proficiency, suggesting a negative impact of implicit aptitude on language outcomes when relied upon independently. It is worth noting that hierarchical regression analysis included explicit aptitude first, adding components of implicit aptitude only after accounting for explicit ones. Thus, this negative effect emerges when the variance attributed to explicit aptitude is removed from the model, indicating that the impact of implicit aptitude turns negative when it becomes the sole basis for prediction. The negative effects are therefore less surprising, particularly given that both implicit and explicit cognitive aptitude jointly contribute to implicit and explicit L2 knowledge, as discussed in Section 9.6 below.

Implicit aptitude was included as a single factor in the multilevel model and was found to positively contribute to automatized explicit knowledge. However, this contribution became evident only after accounting for the mediating effects of proficiency. Specifically, implicit aptitude was shown to play no significant role at lower proficiency levels, becoming beneficial to automatized explicit knowledge only at or above the mean level of speaking proficiency. Implicit aptitude exhibited a similar pattern with implicit and explicit knowledge, albeit at non-significant levels, further highlighting proficiency as a mediator of implicit language aptitude. These findings corroborate results of Linck et al. (2013) and align with theoretical suggestions by Skehan (2016) and Li and DeKeyser (2021), who argue that the automatization occurring at advanced proficiency levels depends on implicit learning abilities. In essence, while proficiency diminishes the effects of explicit aptitude, as seen in Chapter 5, it enhances the role implicit aptitude, making it a significant predictor only at advanced proficiency levels and increasingly so as proficiency grows.

Analysis of structure difficulty as a moderator of implicit aptitude effects indicated that structure difficulty did not significantly influence the impact of implicit aptitude. Reber (1992, 1996) suggests that implicit learning mechanisms, being phylogenetically older, are more stable and thus exhibit less variance compared to the more variable explicit learning mechanisms. This stability likely explains why structure difficulty has minimal effect on implicit aptitude. Supporting this, the models examining the mediating role of proficiency showed that variance for implicit aptitude was limited to a narrow range. Despite an upward trend, with the influence of implicit aptitude increasing as proficiency grew, statistical significance remained confined to this narrow variance range. This indicates that the range of variance ultimately impacts significance and if variance is too limited, statistical significance may not emerge.

9.6. The relationship between implicit and explicit knowledge

Preliminary analysis confirmed that the three L2 tests used in the present study, namely self-paced reading, elicited imitation, and gap-fill, loaded onto two distinct factors: explicit and implicit knowledge. To address the question of the relationship between implicit and explicit knowledge, or the interface issue, the current project adopted a cross-sectional design, examining the relationship between language aptitude and language knowledge to infer causality.

A multilevel model with reading, listening, and speaking proficiency as covariates was created to address the mediating role of proficiency revealed in Chapter 5. Explicit aptitude, consisting of associative memory, phonetic coding ability, and language-analytic ability, predicted implicit knowledge. This predictive relationship was taken as evidence of an explicit-implicit interface, aligning with previous cross-sectional (Suzuki & DeKeyser, 2017) and longitudinal studies (Kim & Godfroid, 2023) that documented an explicit-implicit interface. The results suggest that conscious, explicit language knowledge facilitates the development of implicit representations, consistent with the interface theories of N. Ellis (2005, 2006; 2015), R. Ellis (1994; 2005, 2006), and Hulstijn (2015). N. Ellis's associative-cognitive creed framework (2006) offers the clearest illustration of this interface: in his view, the two types of knowledge interact temporarily during conscious input processing, leaving a lasting impact on implicit cognition and allowing explicit knowledge to shape implicit knowledge over time.

At the same time, an implicit-explicit interface was detected in the model predicting automatized explicit knowledge. Specifically, implicit aptitude, operationalized as implicit sequence learning ability and auditory pattern recognition ability, predicted automatized explicit knowledge. Unlike implicit knowledge, which is unconscious, automatized explicit knowledge is knowledge that has undergone varying levels of automatization and is deployed at a similar speed to implicit knowledge. Thus, the predictive relationship observed in this project suggests that implicitly accrued knowledge can lead to explicit or conscious knowledge. This phenomenon has been documented in both laboratory (Andringa, 2020) and naturalistic settings (Kim & Godfroid, 2023). Implicit-explicit interface implies that rule awareness can emerge from unconscious knowledge, a phenomenon observed in research on language development in children. Indeed, Bialystok's early work (Bialystok, 1994a, 1994b, 2001) introduced a framework accounting for the implicit-explicit direction of the interface. This framework describes language development as a process of increasing explicitness in mental representations, governed by processes of analysis and control. By this reasoning, knowledge can begin as implicit (unconscious knowledge with low control) and, over time, develop into conscious or explicit knowledge with high control as mental representations are reorganized.

9.6.1. Characteristics of the interface

The simultaneous existence of both interfaces can be described as a bidirectional or reciprocal interface (Godfroid, 2023) and has been empirically recorded only once before (Kim & Godfroid, 2023). Bidirectional interface implies a dynamic relationship between the two types of knowledge: implicit or unconscious knowledge can develop with the support of explicit processes, while explicit knowledge can emerge from implicitly learned knowledge.

Comparing findings across studies reveals certain insights. First, participants in the current study were noticeably younger than those in Suzuki and DeKeyser (2017) and Kim and Godfroid (2023), yet all studies found evidence of the interface. This consistency indicates the robustness of the interface, suggesting that it is present across different age groups. Additionally, while participants in the other studies were immersed in the target-language setting without classroom instruction, participants in the current project had limited time in target-language-speaking countries and were primarily exposed to explicit instruction in a classroom setting. This further underscores the robustness of the interface, detected both in immersive settings where learners experience large amounts of input and in classroom settings with limited input and heavily focused on form.

Lastly, the interface was tested for each linguistic structure separately to address Han & Finneran's (2014) proposition that multiple interfaces might exist and that the relationship between implicit and explicit knowledge might vary depending on the difficulty of the linguistic structure. However, the results showed no difference in interface effects despite variations in structure difficulty, providing counter-evidence to the hypothesis that structure difficulty influences the interface between explicit and implicit knowledge.

9.7. Conclusion

9.7.1. Key contributions to the field

The current thesis focused on resolving long-standing questions regarding how language aptitude influences language learning. While it has long been accepted that high language aptitude benefits L2 learning, this thesis presents evidence suggesting that other factors, such as overall proficiency level and the difficulty of the linguistic structures being learned, also contribute to and modify these effects. In light of this new evidence, it becomes essential to consider the roles of L2 proficiency and structure difficulty in any models of language aptitude. Both Skehan's developmental model (2002, 2012, 2016) and Robinson's aptitude complexes model (2005, 2007, 2012) address the dynamic nature of language aptitude, though in distinct ways.

Skehan was the first to formally acknowledge the changing role of aptitude components throughout the language learning process. The results of the current study support Skehan's claim that explicit aptitude components, particularly associative memory, phonetic coding, and language-analytic ability, play a crucial role at beginner and intermediate levels of L2 proficiency. Equally significant is the finding of this project that the influence of these abilities diminishes as learners approach higher proficiency. Although Skehan's model (2016) originally suggested that working memory becomes increasingly important at advanced levels, this is not supported by the current results. Instead, as Li and DeKeyser (2021) argue, cognitive abilities associated with implicit learning begin to play a more prominent role at advanced proficiency levels.

Robinson's aptitude complexes model (2005, 2007, 2012) arguably provides the most comprehensive framework for describing language aptitude. This model not only accounts for the multicomponential nature of both explicit and implicit aptitude but also accommodates varying levels of the cognitive abilities underlying these constructs. By doing so, it allows for the accommodation of positive and negative factor loadings, such as those observed with implicit aptitude in this study. The model also allows for the influence of variables such as proficiency and structure difficulty, which can strengthen or weaken the facilitative effects of language aptitude on learning outcomes, as observed in the current project.

These findings also yield a number of implications for practice. Given that the results support language aptitude as a dynamic concept, teachers can leverage this insight to optimize language instruction. Similarly, adaptive learning technologies, driven by recent advancements in AI, could also benefit from these findings – by carefully considering both the user's proficiency level as well as the relative difficulty of linguistic structures, language learning apps could maximise learning efficiency while taking into account learner's language aptitude.

9.7.2. Limitations and suggestions for future research

The project presented in this thesis has important implications for the role of cognitive individual differences in SLA, providing foundational insights in some areas and key advancements in others, which future studies can build upon. However, like any empirical investigation, this project faced several methodological limitations worth noting.

The first limitation pertains to sample size, one of the most common issues in SLA research. Although this study reached a respectable sample size of 86, it was still insufficiently large to conduct structural equation modeling, which would have been a desirable approach in Articles 2 and 3. The limited sample size prevented examination of the mediating role of proficiency and moderating role of structure difficulty simultaneously for all three structures. As Kline (2016) notes, there is no universal rule for the required sample size for structural equation modeling, as it depends on various factors that influence statistical power. However, a rough estimation suggests that a minimum of N=200 is often required for robust results, with

N=100 as the absolute minimum for simpler models. Similarly, multilevel modeling that incorporated the combined effects of proficiency and structure difficulty, offering a more holistic approach, proved too complex for the current sample size, with models failing to converge. Thus, future studies should include at least 100 participants for simpler analyses, ideally aiming for 200 to achieve more accurate estimations (for more detailed recommendations, see Kline, 2016).

Another limitation involves the serial reaction time task, which is known for yielding low reliability indices, as was observed in this study. Although the reliability scores were consistent with those from previous studies, the low reliability could reflect the challenge of measuring a cognitive ability that operates outside awareness, inevitably introducing noise (Granena, 2020). While this limitation may be intrinsic to the measure, ways to improve the reliability of the serial reaction time task to a recommended level of .7 should be examined.

The issue of low reliability extends to implicit aptitude in general. In this study, LLAMA D and SRT scores loaded together on a factor labelled implicit aptitude, which is empirically coherent (Granena, 2020). However, a stronger case for implicit aptitude could be made by including additional measures. Li and DeKeyser (2021) propose that implicit aptitude encompasses several abilities, such as sensitivity to frequency and conditional probability (gauged by the SRT) and priming or the tendency to be influenced by recent events. While improving the reliability of the SRT task remains essential, adding a measure for priming could provide a more comprehensive assessment of implicit aptitude. Li and DeKeyser (2021) suggest using auditory, semantic, or syntactic priming tasks to capture sensitivity to recent events. Expanding the range of implicit aptitude measures could not only strengthen statistical power but also help explore the structural properties of implicit aptitude, advancing our theoretical understanding – a process that this study has taken initial steps toward.

Similarly, the limited number of measures of implicit knowledge did not allow for SEM. Although it is technically possible to include latent factors composed of single variables, it would be preferable to incorporate multiple measures to create a more reliable implicit knowledge factor. This study originally included two measures of implicit knowledge: a self-paced reading task and a word-monitoring task. However, the word-monitoring task was integrated into an elicited imitation task, a method supported by previous research (Suzuki & DeKeyser, 2015) and validated in pilot testing (see Section 3.4). Despite this, more than 40% of participants in the main study failed to follow instructions, producing unusable data. Many participants, rather than waiting to hear the monitoring word, pressed the button upon seeing it on the screen. Given this outcome, future studies should consider separating the two tasks to reduce task demands and minimize participant confusion.

One of the main contributions of the thesis to SLA is its systematic examination of the mediating role of proficiency on the effects of language aptitude and working memory, with results pinpointing critical tipping points in proficiency. Although the study included a number of proficiency measures, future research should incorporate writing proficiency to allow for the broader generalization of findings. Additionally, while the identified tipping points constitute a valuable and entirely novel contribution, future research could enhance usability by benchmarking proficiency scores against the Common European Framework of Reference (CEFR).

This project also addresses the underexplored role of structure difficulty as a moderating factor, providing insights into this area. However, only three grammatical structures were included, limiting the generalizability of findings. While theoretical and pedagogical reasons justified the selection and difficulty ranking of these structures, including more varied structures and researching L2s other than English could provide stronger estimates and more robust results. This recommendation applies similarly to the examination of the interface issue: while the study provides evidence against structure difficulty moderating the interface, a broader range of structures would allow for further validation of this claim.

Lastly, it should be noted that the results and conclusions are based on the specific sample included in this study, i.e. adolescent learners of English-as-a-foreign-language, aged 15–18, from an instructed learning context. It is well established in cognitive neuroscience that certain cognitive functions, particularly those associated with the prefrontal cortex, continue to undergo significant structural development into the early twenties (Johnson et al., 2009). Therefore, caution is warranted when generalising the findings of this study to adult learners beyond this age range. Similarly, as demonstrated in studies such as Suzuki and DeKeyser (2015), relationships between aptitude and knowledge are often specific to the learning context (e.g. relationship between implicit aptitude and implicit knowledge only for learners who spent considerable time in the target-speaking country). In the current study, implicit knowledge was predicted by explicit aptitude; however, this effect may be specific to instructed learning contexts and may not extend to naturalistic environments. Furthermore, because the participants were drawn from an instructed (classroom) setting where language aptitude is typically associated with the rate of second language learning rather than with long-term or ultimate attainment, the findings should be generalised only to similar contexts. In light of this, future research is encouraged to investigate the roles of implicit and explicit aptitude in areas not addressed by the present thesis.

References

Abrahamsson, N., & Hyltenstam, K. (2009). Age of onset and nativelikeness in a second language: Listener perception versus linguistic scrutiny. *Language Learning*, *59*(2), 249-306. https://doi.org/10.1111/j.1467-9922.2009.00507.x

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

Ackerman, P. L. (2003). Aptitude complexes and trait complexes. *Educational Psychologist*, *38*(2), 85-93. https://doi.org/10.1207/S15326985EP3802_3

Adams, A.-M., & Guillot, K. (2008). Working memory and writing in bilingual students. *International Journal of Applied Linguistics*, *156*(1), 13-28. https://doi.org/10.2143/ITL.156.0.2034417

Aiken, L. S., West, S. G., & Reno, R. R. (1991). *Multiple regression : testing and interpreting interactions*. Sage Publications.

Akamatsu, N. (2008). The effects of training on automatization of word recognition in English as a foreign language. *Applied Psycholinguistics*, *29*(2), 175-193. https://doi.org/10.1017/S0142716408080089

Alptekin, C., & Erçetin, G. (2010). The role of L1 and L2 working memory in literal and inferential comprehension in L2 reading. *Journal of Research in Reading*, *33*(2), 206-219. https://doi.org/10.1111/j.1467-9817.2009.01412.x

Anđel, M., Klampfer, S., Kilani-Schoch, M., Dressler, W. U., & Kovačević, M. (2000). Acquisition of verbs in Croatian, French and Austrian German-an outline of a comparative analysis. *Suvremena lingvistika*, *49*(1), 5-25.

Anderson, J. R. (1982). Acquisition of cognitive skill. *Psychological Review*, *89*(4), 369-406. https://doi.org/10.1037/0033-295X.89.4.369

Anderson, J. R. (2007). *How can the human mind occur in the physical universe?* Oxford University Press.

Andringa, S. (2020). The emergence of awareness in uninstructed L2 learning: A visual world eye tracking study. *Second Language Research*, *36*(3), 335-357. https://doi.org/10.1177/0267658320915502

Andringa, S., & Curcic, M. (2015). How explicit knowledge affects online L2 processing: Evidence from differential object marking acquisition. *Studies in Second Language Acquisition*, *37*(2), 237-268. https://doi.org/10.1017/S0272263115000017

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. https://doi.org/10.1016/j.lindif.2016.06.023

Baddeley, A. D. (1986). *Working memory*. Oxford University Press.

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

Baddeley, A. D. (2015). Working memory in second language learning. In Z. E. Wen, M. B. Mota, & A. McNeill (Eds.), *Working memory in second language acquisition and processing* (pp. 17-28). Multilingual Matters. https://doi.org/10.21832/9781783093595-005

Baddeley, A. D. (2017). Modularity, working memory and language acquisition. *Second Language Research*, *33*(3), 299-311. https://www.jstor.org/stable/26375886

Baddeley, A. D., Gathercole, S., & Papagno, C. (1998). The phonological loop as a language learning device. *Psychological Review*, *105*(1), 158-173. https://doi.org/10.1037/0033-295X.105.1.158

Baddeley, A. D., & Hitch, G. (1974). Working memory. In G. H. Bower (Ed.), *Recent advances in learning and motivation* (pp. 47-89). Academic Press.

Baddeley, A. D., & Logie, R. H. (1999). Working memory: The multicomponent model. In A. Miyake & P. Shah (Eds.), *Models of working memory: Mechanisms of active maintenance and executive control* (pp. 28-61). Cambridge University Press.

Baddeley, A. D., Thomson, N., & Buchanan, M. (1975). Word length and the structure of short-term memory. *Journal of Verbal Learning & Verbal Behavior*, *14*(6), 575-589. https://doi.org/10.1016/S0022-5371(75)80045-4

Bagarić, V. (2001). Language awareness and the learning of English and German as foreign languages. *Strani jezici*, *30*(3), 107-121.

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

Bialystok, E. (1994a). Analysis and control in the development of second language proficiency. *Studies in Second Language Acquisition*, *16*(2), 157-168. https://doi.org/10.1017/S0272263100012857

Bialystok, E. (1994b). Representation and ways of knowing: Three issues in second language acquisition. In N. C. Ellis (Ed.), *Implicit and explicit learning of languages* (pp. 549-569). Academic Press.

Bialystok, E. (2001). *Bilingualism in development : language, literacy, and cognition*. Cambridge University Press.

Biedroń, A., Véliz-Campos, M., & Zychowicz, K. (2022). Longitudinal effects of phonological short-term memory and working memory capacity on L2 grammar knowledge. *Advances in Cognitive Psychology*, *18*(2), 113-121. https://doi.org/10.5709/acp-0353-1 bin Abdullah, A. T. H. (2013). Error analysis on the use of the simple past tense and the simple past tense in writing essays among TESL college students. *International Journal of Education and Research*, *1*(12), 1-12.

Bley-Vroman, R. W., & Chaudron, C. (1994). Elicited imitation as a measure of second-language competence. In E. Tarone, S. M. Gass, & A. D. Cohen (Eds.), *Research methodology in second-language acquisition* (pp. 245-261). L. Erlbaum.

Bogaerts, L., Siegelman, N., & Frost, R. (2020). Statistical learning and language impairments: Toward more precise theoretical accounts. *Perspectives on Psychological Science*, *16*(2), 319-337. https://doi.org/10.1177/1745691620953082

Bokander, L., & Bylund, E. (2020). Probing the internal validity of the LLAMA language aptitude tests. *Language Learning*, *70*(1), 11-47. https://doi.org/10.1111/lang.12368

Brown, T. A. (2015). *Confirmatory factor analysis for applied research* (Second edition. ed.). Guilford Publications.

Buffington, J., Demos, A. P., & Morgan-Short, K. (2021). The reliability and validity of procedural memory assessments used in second language acquisition research. *Studies in Second Language Acquisition*, *43*(3), 635-662. https://doi.org/10.1017/S0272263121000127

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

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

Carroll, J. B. (1990). Cognitive abilities in foreign language aptitude: Then and now. In T. S. Parry & C. W. Stansfield (Eds.), *Language aptitude reconsidered* (pp. 11-27). Prentice Hall Regents.

Carroll, J. B. (1993). *Human cognitive abilities: A survey of factor-analytic studies*. Cambridge University Press https://doi.org/10.1017/CBO9780511571312

Carroll, J. B., & Sapon, S. M. (1959). Modern language aptitude test.

Cheung, H. (1996). Nonword span as a unique predictor of second-language vocabulary language. *Developmental Psychology*, *32*(5), 867–873. https://doi.org/10.1037/0012-1649.32.5.867

Christiansen, M. H. (2019). Implicit statistical learning: A tale of two literatures. *Topics in Cognitive Science*, *11*(3), 468-481. https://doi.org/10.1111/tops.12332

Clahsen, H., & Felser, C. (2006). Grammatical processing in language learners. *Applied Psycholinguistics*, *27*(1), 3-42. https://doi.org/10.1017/S0142716406060024

Cleeremans, A. (2008). Consciousness: The radical plasticity thesis. In R. Banerjee & B. K. Chakrabarti (Eds.), *Progress in Brain Research* (Vol. 168, pp. 19-33). Elsevier. https://doi.org/10.1016/S0079-6123(07)68003-0

Cleeremans, A. (2011). The radical plasticity thesis: How the brain learns to be conscious. *Frontiers in Psychology*, *2*(86). https://doi.org/10.3389/fpsyg.2011.00086

Cogan, G. B., Iyer, A., Melloni, L., Thesen, T., Friedman, D., Doyle, W., Devinsky, O., & Pesaran, B. (2017). Manipulating stored phonological input during verbal working memory. *Nature Neuroscience*, *20*(2), 279-286. https://doi.org/10.1038/nn.4459

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Lawrence Erlbaum Associates.

Collier, G. L., & Logan, G. (2000). Modality differences in short-term memory for rhythms. *Memory & Cognition*, *28*(4), 529-538. https://doi.org/10.3758/BF03201243

Collins, L., Trofimovich, P., White, J., Cardoso, W., & Horst, M. (2009). Some input on the easy/difficult grammar question: An empirical study. *The Modern Language Journal*, *93*(3), 336-353. https://doi.org/10.1111/j.1540-4781.2009.00894.x

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

Conway, C. M., & Christiansen, M. H. (2006). Statistical learning within and between modalities: Pitting abstract against stimulus-specific representations. *Psychological Science*, *17*(10), 905-912. https://doi.org/10.1111/j.1467-9280.2006.01801.x

Conway, C. M., Pisoni, D. B., & Kronenberger, W. G. (2009). The importance of sound for cognitive sequencing abilities: The auditory scaffolding hypothesis. *Current Directions in Psychological Science*, *18*(5), 275-279. https://doi.org/10.1111/j.1467-8721.2009.01651.x

Cowan, N. (1999). An embedded-processes model of working memory. In A. Miyake & P. Shah (Eds.), *Models of working memory: Mechanisms of active maintenance and executive control* (pp. 62-101). Cambridge University Press.

Cowan, N. (2005). *Working memory capacity*. Psychology Press. https://doi.org/10.4324/9780203342398

Croft, W. (1991). *Syntactic categories and grammatical relations: The cognitive organization of information*. University of Chicago Press.

Cunnings, I., & Finlayson, I. (2015). Mixed effects modeling and longitudinal data analysis. In L. Plonsky (Ed.), *Advancing quantitative methods in second language research*. Routledge. https://doi.org/10.4324/9781315870908

Čupić, I., & Klanjčić, M. (2015). Making students more active in learning the passive. https://doi.org/10.14706/jfltal151312

Curcic, M., Andringa, S., & Kuiken, F. (2019). The role of awareness and cognitive aptitudes in L2 predictive language processing. *Language Learning*, *69*(S1), 42-71. https://doi.org/10.1111/lang.12321

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

DeKeyser, R. M. (1998). *Beyond focus on form: Cognitive perspectives on learning and practicing second language grammar*. In J. Williams & C. Doughty (Eds.), *Focus on form in classroom second language acquisition* (pp. 42-63). Cambridge University Press.

DeKeyser, R. M. (2003). Implicit and explicit learning. In C. J. Doughty & M. H. Long (Eds.), *The handbook of second language acquisition* (pp. 313-348). Blackwell Publishing Ltd.

DeKeyser, R. M. (2005). What makes learning second-language grammar difficult? A review of issues. *Language Learning*, *55*(S1), 1-25. https://doi.org/10.1111/j.0023-8333.2005.00294.x

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

DeKeyser, R. M. (2017). Knowledge and skill in ISLA. In S. Loewen & M. Sato (Eds.), *The Routledge Handbook of Instructed Second Language Acquisition* (1st ed., pp. 15-32). Routledge.

DeKeyser, R. M. (2020). Skill acquisition theory. In B. VanPatten, G. D. Keating, & S. Wulff (Eds.), *Theories in second language acquisition: An introduction* (3rd ed., pp. 106-124). Routledge, Taylor & Francis Group. https://doi.org/10.4324/9780429503986-5

DeKeyser, R. M. (2021). *Aptitude-treatment interaction in second language learning*. John Benjamins Publishing Company.

DeKeyser, R. M., & Larson-Hall, J. (2005). What does the critical period really mean? In J. F. Kroll & A. M. B. d. Groot (Eds.), *Handbook of bilingualism: Psycholinguistic approaches* (pp. 88-108). Oxford University Press.

DeKeyser, R. M., & Li, S. (2021). Epilogue: Evidence for the validity of implicit aptitude and the need for construct validation. *Studies in Second Language Acquisition*, *43*(3), 692-697. https://doi.org/10.1017/S0272263121000140

Dienes, Z. (1992). Connectionist and memory-array models of artificial grammar learning. *Cognitive Science*, *16*(1), 41-79. https://doi.org/10.1016/0364-0213(92)90017-O

Dienes, Z. (2008). Subjective measures of unconscious knowledge. *Progress in brain research*, *168*, 49-64. https://doi.org/10.1016/S0079-6123(07)68005-4

Dienes, Z., & Scott, R. (2005). Measuring unconscious knowledge: distinguishing structural knowledge and judgment knowledge. *Psychological Research*, *69*(5-6), 338–351. https://doi.org/10.1007/s00426-004-0208-3

Dienes, Z., Williams, J. N., & Rebuschat, P. (2012). Conscious versus unconscious learning of structure. In P. Rebuschat & J. N. Williams (Eds.), *Statistical learning and language acquisition* (1st ed., Vol. 1, pp. 337-364). De Gruyter, Inc. https://doi.org/10.1515/9781934078242.337

Dörnyei, Z. (2005). *The psychology of the language learner: Individual differences in second language acquisition* (1st ed.). Lawrence Erlbaum Associates, Inc.

Dörnyei, Z., & Skehan, P. (2003). Individual differences in second language learning. In C. J. Doughty & M. H. Long (Eds.), *The Handbook of Second Language Acquisition*. Blackwell Publishing Ltd.

Draganski, B., Gaser, C., Busch, V., Schuierer, G., Bogdahn, U., & May, A. (2004). Changes in grey matter induced by training. *Nature*, *427*(6972), 311-312. https://doi.org/10.1038/427311a

Education First. (2024). *The world's largest ranking of countries and regions by English skills*. Retrieved 25 May 2025 from https://www.ef.com/wwen/epi/

Ellis, N. (1996). Sequencing in SLA: Phonological memory, chunking, and points of order. *Studies in Second Language Acquisition*, *18*(1), 91-126. https://doi.org/10.1017/S0272263100014698

Ellis, N. (2005). *At the interface: Dynamic interactions of explicit and implicit language knowledge*. *Studies in Second Language Acquisition*, *27*(2), 305-352. https://doi.org/10.1017/S027226310505014X

Ellis, N. (2006). Cognitive perspectives on SLA: The associative-cognitive CREED. *AILA Review*, *19*(1), 100-121. https://doi.org/10.1075/aila.19.08ell

Ellis, N. (2011). Implicit and explicit SLA and their interface. In C. Sanz & R. P. Leow (Eds.), *Implicit and explicit language learning* (pp. 35-47). Georgetown University Press.

Ellis, N. (2015). Implicit AND explicit language learning: Their dynamic interface and complexity. In P. Rebuschat (Ed.), *Implicit and explicit learning of languages* (pp. 3-24). John Benjamins Publishing Company. https://doi.org/10.1075/sibil.48.01ell

Ellis, N., & Wulff, S. (2020). Usage-based approaches to L2 acquisition. In B. VanPatten, G. D. Keating, & S. Wulff (Eds.), *Theories in second language acquisition: An introduction* (3rd ed., pp. 63-82). Routledge, Taylor & Francis Group. http://ebookcentral.proquest.com/lib/universityofessex-ebooks/detail.action?docID=6120993

Ellis, R. (1994). A theory of instructed second language acquisition. In N. Ellis (Ed.), *Implicit and explicit learning of languages* (pp. *79–114*). Academic Press.

Ellis, R. (2001). Introduction: Investigating form-focused instruction. *Language Learning*, *51*(s1), 1-46. https://doi.org/10.1111/j.1467-1770.2001.tb00013.x

Ellis, R. (2002). Does form-focused instruction affect the acquisition of implicit knowledge?: A review of the research. *Studies in Second Language Acquisition*, *24*(2), 223-236. https://doi.org/10.1017/S0272263102002073

Ellis, R. (2004). The definition and measurement of L2 explicit knowledge. *Language Learning*, *54*(2), 227-275. https://doi.org/10.1111/j.1467-9922.2004.00255.x

Ellis, R. (2005). Measuring implicit and explicit knowledge of a second language: A psychometric study. *Studies in Second Language Acquisition*, *27*(2), 141-172. https://doi.org/10.1017/S0272263105050096

Ellis, R. (2006). Modelling learning difficulty and second language proficiency: The differential contributions of implicit and explicit knowledge. *Applied Linguistics*, *27*(3), 431-463. https://doi.org/10.1093/applin/aml022

Ellis, R. (2009a). Investigating learning difficulty in terms of implicit and explicit knowledge. In R. Ellis, S. Loewen, C. Elder, R. Erlam, J. Philp, & H. Reinders (Eds.), *Implicit and explicit knowledge in second language learning, testing and teaching* (pp. 143-166). Multilingual Matters. https://doi.org/10.21832/9781847691767-008

Ellis, R. (2009b). Measuring implicit and explicit knowledge of a second language. In R. Ellis, S. Loewen, C. Elder, R. Erlam, J. Philp, & H. Reinders (Eds.), *Implicit and explicit knowledge in second language learning, testing and teaching* (pp. 31-64). Multilingual Matters. https://doi.org/10.21832/9781847691767-004

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.), *The psychology of learning and motivation* (Vol. 44, pp. 145-199).

Engle, R. W., Nations, J. K., & Cantor, J. (1990). Is "working memory capacity" just another name for word knowledge? *Journal of Educational Psychology*, *82*(4), 799-804. https://doi.org/10.1037/0022-0663.82.4.799

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

Erlam, R. (2006). Elicited imitation as a measure of L2 implicit knowledge: An empirical validation study. *Applied Linguistics*, *27*(3), 464-491. https://doi.org/10.1093/applin/aml001

Evans, J. S. B. T., & Over, D. E. (1996). *Rationality and reasoning* (1st ed.). Psychology Press. https://doi.org/10.4324/9780203027677

Felser, C., & Roberts, L. (2007). Processing wh-dependencies in a second language: A cross-modal priming study. *Second Language Research*, *23*(1), 9-36. https://doi.org/10.1177/0267658307071600

Field, A. P. (2018). *Discovering statistics using IBM SPSS statistics* (5th ed.). SAGE.

Foerde, K., Knowlton, B. J., & Poldrack, R. A. (2006). Modulation of competing memory systems by distraction. *Proceedings of the National Academy of Sciences*, *103*(31), 11778-11783. https://doi.org/10.1073/pnas.0602659103

Foster, P., Bolibaugh, C., & Kotula, A. (2013). Knowledge of nativelike selections in a L2: The influence of exposure, memory, age of onset, and motivation in foreign language and immersion settings. *Studies in Second Language Acquisition*, *36*(1), 101-132. https://doi.org/10.1017/S0272263113000624

Foster, P., Tonkyn, A., & Wigglesworth, G. (2000). Measuring spoken language: A unit for all reasons. *Applied Linguistics*, *21*(3), 354-375. https://doi.org/10.1093/applin/21.3.354

Freed, B. (2000). Is fluency, like beauty, in the eyes (and ears) of the beholder? In H. Riggenbach (Ed.), *Perspectives on fluency* (pp. 243-265). University of Michigan Press. https://doi.org/10.3998/mpub.16109

French, L. M., & O'Brien, I. (2008). Phonological memory and children's second language grammar learning. *Applied Psycholinguistics*, *29*(3), 463-487. https://doi.org/10.1017/S0142716408080211

Frost, R., Siegelman, N., Narkiss, A., & Afek, L. (2013). What predicts successful literacy acquisition in a second language? *Psychological Science*, *24*(7), 1243-1252. https://doi.org/10.1177/0956797612472207

Gathercole, S. E. (1995). Is nonword repetition a test of phonological memory or long-term knowledge? It all depends on the nonwords. *Memory & Cognition*, *23*(1), 83-94. https://doi.org/10.3758/BF03210559

Gathercole, S. E. (2006). Nonword repetition and word learning: The nature of the relationship. *Applied Psycholinguistics*, *27*(4), 513-543. https://doi.org/10.1017/S0142716406060383

Gevins, A., Cutillo, B., & Smith, M. E. (1995). Regional modulation of high resolution evoked potentials during verbal and non-verbal matching tasks. *Electroencephalography and Clinical Neurophysiology*, *94*(2), 129-147. https://doi.org/10.1016/0013-4694(94)00261-I

Godfroid, A. (2016). The effects of implicit instruction on implicit and explicit knowledge development. *Studies in Second Language Acquisition*, *38*(2), 177-215. https://doi.org/10.1017/S0272263115000388

Godfroid, A. (2023). Hypotheses about the interface between explicit and implicit knowledge in second language acquisition. In A. Godfroid & H. Hopp (Eds.), *The Routledge*

Handbook of Second Language Acquisition and Psycholinguistics (pp. 294-307). Routledge, Taylor & Francis Group.

Godfroid, A., & Kim, K. M. (2021). The contributions of implicit-statistical learning aptitude to implicit second-language knowledge. *Studies in Second Language Acquisition*, *43*(3), 606-634. https://doi.org/10.1017/S0272263121000085

Granena, G. (2012). *Age differences and cognitive aptitudes for implicit and explicit learning in ultimate second language attainment*. University of Maryland, College Park.

Granena, G. (2013a). Cognitive aptitudes for second language learning and the LLAMA Language Aptitude Test. In G. Granena & M. H. Long (Eds.), *Sensitive Periods, Language Aptitude, and Ultimate L2 Attainment* (pp. 105-130). John Benjamins Publishing Company. https://doi.org/10.1075/lllt.35.04gra

Granena, G. (2013b). Individual differences in sequence learning ability and second language acquisition in early childhood and adulthood. *Language Learning*, *63*(4), 665-703. https://doi.org/10.1111/lang.12018

Granena, G. (2013c). Reexamining the robustness of aptitude in second language acquisition. In G. Granena & M. H. Long (Eds.), *Sensitive Periods, Language Aptitude, and Ultimate L2 Attainment*. John Benjamins Publishing Company. https://doi.org/https://doi.org/10.1075/lllt.35.07gra

Granena, G. (2016). Cognitive aptitudes for implicit and explicit learning and information-processing styles: An individual differences study. *Applied Psycholinguistics*, *37*(3), 577-600. https://doi.org/10.1017/S0142716415000120

Granena, G. (2019). Cognitive aptitudes and L2 speaking proficiency: Links between LLAMA and Hi-LAB. *Studies in Second Language Acquisition*, *41*(2), 313-336. https://doi.org/10.1017/S0272263118000256

Granena, G. (2020). *Implicit language aptitude*. Cambridge University Press. https://doi.org/10.1017/9781108625616

Granena, G., & Yilmaz, Y. (2019). Corrective feedback and the role of implicit sequence-learning ability in L2 online performance. *Language Learning*, *69*(S1), 127-156. https://doi.org/10.1111/lang.12319

Grey, S., Cox, J. G., Serafini, E. J., & Sanz, C. (2015). The role of individual differences in the study abroad context: Cognitive capacity and language development during short-term intensive language exposure. *The Modern Language Journal*, *99*(1), 137-157. https://doi.org/10.1111/modl.12190

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. https://doi.org/10.1111/0026-7902.00076

Guiraud, P. (1954). Les charact`eres statistiques du vocabulaire. In *Essai de m´ethodologie*. Presses Universitaires de France.

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

Han, Z., & Finneran, R. (2014). Re-engaging the interface debate: Strong, weak, none, or all? *International Journal of Applied Linguistics*, *24*(3), 370-389. https://doi.org/10.1111/ijal.12034

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

Harrison, T. L., Shipstead, Z., Hicks, K. L., Hambrick, D. Z., Redick, T. S., & Engle, R. W. (2013). Working memory training may increase working memory capacity but not fluid intelligence. *Psychological Science*, *24*(12), 2409-2419. https://doi.org/10.1177/0956797613492984

Hawkins, J. A. (2015). *Definiteness and indefiniteness : A study in reference and grammaticality prediction*. Routledge. https://doi.org/10.4324/9781315687919

Housen, A., & Simoens, H. (2016). Introduction: Cognitive perspectives on difficulty and complexity in L2 acquisition. *Studies in Second Language Acquisition*, *38*(2), 163-175. https://doi.org/10.1017/S0272263116000176

Howard, J. H., & Howard, D. V. (1997). Age differences in implicit learning of higher order dependencies in serial patterns. *Psychology and aging*, *12*(4), 634–656. https://doi.org/10.1037//0882-7974.12.4.634

Hulstijn, J. H. (1990). A comparison between the information-processing and the analysis/control approaches to language learning. *Applied Linguistics*, *11*(1), 30-45. https://doi.org/10.1093/applin/11.1.30

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

Hulstijn, J. H. (2015). Explaining phenomena of first and second language acquisition with the constructs of implicit and explicit learning: The virtues and pitfalls of a two-system view. In P. Rebuschat (Ed.), *Implicit and explicit learning of languages* (pp. 25-46). John Benjamins Publishing Company. https://doi.org/10.1075/sibil.48.02hul

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

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

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

IBM Corp. (2020). *IBM SPSS Statistics for Macintosh*. In (Version 27.0.) IBM Corp.

IBM Corp. (2023). *IBM SPSS Statistics for Macintosh*. In (Version 29.0.) IBM Corp.

Iizuka, T., & DeKeyser, R. M. (2023). Scrutinizing LLAMA D as a measure of implicit learning aptitude. *Studies in Second Language Acquisition*, 1-23. https://doi.org/10.1017/S0272263122000559

Jeon, E. H. (2015). Multiple regression. In L. Plonsky (Ed.), *Advancing Quantitative Methods in Second Language Research*. Routledge. https://doi.org/10.4324/9781315870908

Jiang, N. (2004). Morphological insensitivity in second language processing. *Applied Psycholinguistics*, *25*(4), 603-634. https://doi.org/10.1017/S0142716404001298

Jiang, N. (2007). Selective integration of linguistic knowledge in adult second language learning. *Language Learning*, *57*(1), 1-33. https://doi.org/10.1111/j.1467-9922.2007.00397.x

Jiang, N., Novokshanova, E., Masuda, K., & Wang, X. (2011). Morphological congruency and the acquisition of L2 morphemes. *Language Learning*, *61*(3), 940-967. https://doi.org/10.1111/j.1467-9922.2010.00627.x

Jiménez, L., & Vázquez, G. A. (2005). Sequence learning under dual-task conditions: Alternatives to a resource-based account. *Psychological Research*, *69*(5), 352-368. https://doi.org/10.1007/s00426-004-0210-9

Johnson, P. O., & Neyman, J. (1936). Tests of certain linear hypotheses and their application to some educational problems. *Statistical Research Memoirs*, *1*, 57-93.

Johnson, S. B., Blum, R. W., & Giedd, J. N. (2009). Adolescent maturity and the brain: The promise and pitfalls of neuroscience research in adolescent health policy. *The Journal of adolescent health: Official publication of the Society for Adolescent Medicine*, *45*(3), 216-221. https://doi.org/10.1016/j.jadohealth.2009.05.016

Juffs, A. (2004). Representation, processing and working memory in a second language. *Transactions of the Philological Society*, *102*(2), 199-225. https://doi.org/10.1111/j.0079-1636.2004.00135.x

Juffs, A. (2005). The influence of first language on the processing of wh-movement in English as a second language. *Second Language Research*, *21*(2), 121-151. https://doi.org/10.1191/0267658305sr255oa

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

Just, M. A., & Carpenter, P. A. (1992). A capacity theory of comprehension: Individual differences in working memory. *Psychological Review*, *99*(1), 122-149. https://doi.org/10.1037/0033-295X.99.1.122

Kaller, C. P., Rahm, B., Köstering, L., & Unterrainer, J. M. (2011). Reviewing the impact of problem structure on planning: A software tool for analyzing tower tasks. *Behavioural Brain Research*, *216*(1), 1-8. https://doi.org/10.1016/j.bbr.2010.07.029

Kaller, C. P., Unterrainer, J. M., & Stahl, C. (2012). Assessing planning ability with the Tower of London task: Psychometric properties of a structurally balanced problem set. *Psychological Assessment*, *24*(1), 46–53. https://doi.org/10.1037/a0025174

Kalra, P. B., Gabrieli, J. D. E., & Finn, A. S. (2019). Evidence of stable individual differences in implicit learning. *Cognition*, *190*, 199-211. https://doi.org/10.1016/j.cognition.2019.05.007

Karmiloff-Smith, A. (1992). *Beyond modularity: A developmental perspective on cognitive science*. MIT Press.

Kaufman, S. B., DeYoung, C. G., Gray, J. R., Jiménez, L., Brown, J., & Mackintosh, N. (2010). Implicit learning as an ability. *Cognition*, *116*(3), 321-340. https://doi.org/10.1016/j.cognition.2010.05.011

Kempe, V., & Brooks, P. J. (2008). Second language learning of complex inflectional systems. *Language Learning*, *58*(4), 703-746. https://doi.org/10.1111/j.1467-9922.2008.00477.x

Kempe, V., Brooks, P. J., & Kharkhurin, A. (2010). Cognitive predictors of generalization of Russian grammatical gender categories. *Language Learning*, *60*(1), 127-153. https://doi.org/10.1111/j.1467-9922.2009.00553.x

Kilborn, H., & Moss, K. (1996). Word monitoring. *Language and Cognitive Processes*, *11*(6), 689-694. https://doi.org/10.1080/016909696387105

Kim, K. M., & Godfroid, A. (2023). The interface of explicit and implicit second-language knowledge: A longitudinal study. *Bilingualism: Language and Cognition*, *26*(4), 709-723. https://doi.org/10.1017/S1366728922000773

Kline, R. B. (2016). *Principles and practice of structural equation modeling*. The Guildford Press.

Knowlton, B. J., Mangels, J. A., & Squire, L. R. (1996). A neostriatal habit learning system in humans. *Science*, *273*(5280), 1399-1402. https://doi.org/10.1126/science.273.5280.1399

Knowlton, B. J., Squire, L. R., & Gluck, M. A. (1994). Probabilistic classification learning in amnesia. *Learning & memory*, *1*(2), 106–120.

Kormos, J., & Sáfár, A. (2008). Phonological short-term memory, working memory and foreign language performance in intensive language learning. *Bilingualism: Language and Cognition*, *11*(2), 261-271. https://doi.org/10.1017/S1366728908003416

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

Krashen, S. D. (1982). *Principles and practice in second language acquisition*. Pergamon.

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

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software*, *82*(13), 1 - 26. https://doi.org/10.18637/jss.v082.i13

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. https://doi.org/10.1111/j.1467-9922.2007.00408.x

Leow, R. P. (2015). *Explicit learning in the L2 classroom: A student-centered approach* (1 ed.). Routledge. https://doi.org/10.4324/9781315887074

Li, S. (2013). The interactions between the effects of implicit and explicit feedback and individual differences in language analytic ability and working memory. *The Modern Language Journal*, *97*(3), 634-654. https://doi.org/10.1111/j.1540-4781.2013.12030.x

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

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

Li, S. (2022). Explicit and implicit language aptitudes. In S. Li, P. Hiver, & M. Papi (Eds.), *The Routledge handbook of second language acquisition and individual differences*. Routledge. https://doi.org/10.4324/9781003270546

Li, S., & DeKeyser, R. M. (2021). Implicit language aptitude: Conceptualizing the construct, validating the measures, and examining the evidence: Introduction to the special issue.

Studies in Second Language Acquisition, *43*(3), 473-497. https://doi.org/10.1017/S0272263121000024

Li, S., Ellis, R., & Zhu, Y. (2019). The associations between cognitive ability and L2 development under five different instructional conditions. *Applied Psycholinguistics*, *40*(3), 693-722. https://doi.org/10.1017/S0142716418000796

Li, S., & Qian, J. (2021). Exploring syntactic priming as a measure of implicit language aptitude. *Studies in Second Language Acquisition*, *43*(3), 574-605. https://doi.org/10.1017/S0272263120000698

Linck, J. A. (2016). Analyzing individual differences in second language research: The benefits of mixed effects models. In G. Granena, D. O. Jackson, & Y. Yilmaz (Eds.), *Cognitive individual differences in second language processing and acquisition* (pp. 105-128). https://doi.org/10.1075/bpa.3.06lin

Linck, J. A., Hughes, M. M., Campbell, S. G., Silbert, N. H., Tare, M., Jackson, S. R., Smith, B. K., Bunting, M. F., & Doughty, C. J. (2013). Hi-LAB: A new measure of aptitude for high-level language proficiency. *Language Learning*, *63*(3), 530-566. https://doi.org/10.1111/lang.12011

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

Linck, J. A., & Weiss, D. J. (2011). Working memory predicts the acquisition of explicit L2 knowledge. In C. Sanz & R. P. Leow (Eds.), *Implicit and Explicit Language Learning* (pp. 101-114). Georgetown University Press.

Liu, D., & Gleason, J. L. (2002). Acquisition of the article the by nonnative speakers of English: An analysis of four nongeneric uses. *Studies in Second Language Acquisition*, *24*(1), 1-26. https://doi.org/10.1017/S0272263102001018

Long, J. A. (2024). *interactions: Comprehensive, user-friendly toolkit for probing interactions*. In https://cran.r-project.org/package=interactions

Long, M. H. (1996). The role of the linguistic environment in second language acquisition. In W. C. Ritchie & T. K. Bhatia (Eds.), *Handbook of second language acquisition* (pp. *413–468*). Academic Press.

Long, M. H. (2005). Problems with supposed counter-evidence to the Critical Period Hypothesis. *43*(4), 287-317. https://doi.org/10.1515/iral.2005.43.4.287 (International Review of Applied Linguistics in Language Teaching)

Long, M. H. (2015). *Second language acquisition and task-based language teaching* (1st ed.). Wiley Blackwell.

MacIntyre, P. D. (2002). Motivation, anxiety and emotion in second language acquisition. In P. Robinson (Ed.), *Individual differences and instructed language learning* (pp. 45-68). John Benjamins Publishing Company. https://doi.org/10.1075/lllt.2.05mac

Mackey, A., Adams, R., Stafford, C., & Winke, P. (2010). Exploring the relationship between modified output and working memory capacity. *Language Learning*, *60*(3), 501-533. https://doi.org/10.1111/j.1467-9922.2010.00565.x

Mackey, A., & Gass, S. M. (2022). *Second language research: Methodology and design* (3rd ed.). Routledge. https://doi.org/10.4324/9781003188414

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

MacWhinney, B. (2000). *The CHILDES project: Tools for analyzing talk: Transcription format and programs, Vol. 1, 3rd ed*. Lawrence Erlbaum Associates Publishers.

Maguire, E. A., Gadian, D. G., Johnsrude, I. S., Good, C. D., Ashburner, J., Frackowiak, R. S. J., & Frith, C. D. (2000). Navigation-related structural change in the hippocampi of taxi drivers. *Proceedings of the National Academy of Sciences*, *97*(8), 4398-4403. https://doi.org/10.1073/pnas.070039597

Maie, R., & DeKeyser, R. M. (2020). Conflicting evidence of explicit and implicit knowledge from objective and subjective measures. *Studies in Second Language Acquisition*, *42*(2), 359-382. https://doi.org/10.1017/S0272263119000615

Marsden, E., Thompson, S., & Plonsky, L. (2018). A methodological synthesis of self-paced reading in second language research. *Applied Psycholinguistics*, *39*(5), 861-904. https://doi.org/10.1017/S0142716418000036

Marslen-Wilson, W., & Tyler, L. K. (1975). Processing structure of sentence perception. *Nature*, *257*(5529), 784-786. https://doi.org/10.1038/257784a0

Marslen-Wilson, W., & Tyler, L. K. (1980). The temporal structure of spoken language understanding. *Cognition*, *8*(1), 1-71. https://doi.org/10.1016/0010-0277(80)90015-3

Masgoret, A. M., & Gardner, R. C. (2003). Attitudes, motivation, and second language learning: A meta–analysis of studies conducted by Gardner and associates. *Language Learning*, *53*(1), 123-163. https://doi.org/10.1111/1467-9922.00212

Masoura, E. V., & Gathercole, S. E. (2005). Contrasting contributions of phonological short-term memory and long-term knowledge to vocabulary learning in a foreign language. *Memory*, *13*(3-4), 422-429. https://doi.org/10.1080/09658210344000323

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

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

Medved Krajnović, M. (2007). How well do Croatian learners speak English? *Metodika*, *8*(1), 182-189. https://hrcak.srce.hr/26951

Michel, M. C., Kuiken, F., & Vedder, I. (2007). The influence of complexity in monologic versus dialogic tasks in Dutch L2. *International Review of Applied Linguistics in Language Teaching*, *45*(3), 241-259. https://doi.org/10.1515/iral.2007.011

Miller, G. A., Galanter, E., & Pribram, K. H. (1960). *Plans and the structure of behavior*. Henry Holt and Co. https://doi.org/10.1037/10039-000

Misyak, J. B., & Christiansen, M. H. (2012). Statistical learning and language: An individual differences study. *Language Learning*, *62*(1), 302-331. https://doi.org/10.1111/j.1467-9922.2010.00626.x

Misyak, J. B., Christiansen, M. H., & Tomblin, J. B. (2010). On-line individual differences in statistical learning predict language processing. *Frontiers in Psychology*, *1*(31). https://doi.org/10.3389/fpsyg.2010.00031

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

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

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

Miyake, A., & Shah, P. (1999). *Models of working memory: Mechanisms of active maintenance and executive control*. Cambridge University Press.

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

Morgan-Short, K., Sanz, C., Steinhauer, K., & Ullman, M. T. (2010). Second language acquisition of gender agreement in explicit and implicit training conditions: An event-related potential study. *Language Learning*, *60*(1), 154-193. https://doi.org/10.1111/j.1467-9922.2009.00554.x

Morgan-Short, K., Steinhauer, K., Sanz, C., & Ullman, M. T. (2012). Explicit and implicit second language training differentially affect the achievement of native-like brain activation patterns. *Journal of Cognitive Neuroscience*, *24*(4), 933-947. https://doi.org/10.1162/jocn_a_00119

Nissen, M. J., & Bullemer, P. (1987). Attentional requirements of learning: Evidence from performance measures. *Cognitive Psychology*, *19*(1), 1-32. https://doi.org/10.1016/0010-0285(87)90002-8

Norris, J. M., & Ortega, L. (2000). Effectiveness of L2 instruction: A research synthesis and quantitative meta-analysis. *Language Learning*, *50*(3), 417-528. https://doi.org/10.1111/0023-8333.00136

Norris, J. M., & Ortega, L. (2006). *Synthesizing research on language learning and teaching*. John Benjamins Publishing Company.

Novick, J. M., Hussey, E., Teubner-Rhodes, S., Harbison, J. I., & Bunting, M. F. (2013). Clearing the garden-path: Improving sentence processing through cognitive control training. *Language, Cognition and Neuroscience*, *29*(2), 186-217. https://doi.org/10.1080/01690965.2012.758297

O'Brien, I., Segalowitz, N., Collentine, J. O. E., & Freed, B. (2006). Phonological memory and lexical, narrative, and grammatical skills in second language oral production by adult learners. *Applied Psycholinguistics*, *27*(3), 377-402. https://doi.org/10.1017/S0142716406060322

O'Grady, W., Archibald, J., Aronoff, J., & Rees-Miller, J. (2001). *Contemporary linguistics: An introduction* (4th ed.). St. Martin's Press.

Oberauer, K., Süß, H.-M., Wilhelm, O., & Wittman, W. W. (2003). The multiple faces of working memory: Storage, processing, supervision, and coordination. *Intelligence*, *31*(2), 167-193. https://doi.org/10.1016/S0160-2896(02)00115-0

Okura, E., & Lonsdale, D. (2012). Working memory's meager involvement in sentence repetition tests. 34th Annual Conference of the Cognitive Science Society, Austin, TX.

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. https://doi.org/10.3758/BF03330466

Oxford University Press. (n.d.). *Oxford Placement Test*. https://elt.oup.com/feature/global/oxford-online-placement/?cc=gb&selLanguage=en

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

Paradis, M. (1994). Neurolinguistic aspects of implicit and explicit memory: Implications for bilingualism. In N. Ellis (Ed.), *Implicit and explicit learning of languages* (pp. 393–419). Academic Press.

Paradis, M. (2004). *A neurolinguistic theory of bilingualism*. J. Benjamin. https://doi.org/10.1075/sibil.18

Paradis, M. (2009). *Declarative and procedural determinants of second languages*. John Benjamins. https://doi.org/10.1075/sibil.40

Pavlekovic, R., & Roehr-Brackin, K. (2024). Aptitude for explicit and implicit learning. *Journal of the European Second Language Association*. https://doi.org/10.22599/jesla.117

Pavlekovic, R., & Roehr-Brackin, K. (under review). Finding the tipping point: Proficiency as a mediator of aptitude in L2 learning. *Studies in Second Language Acquisition*.

Pawlak, M., & Biedroń, A. (2021). Working memory as a factor mediating explicit and implicit knowledge of English grammar. *Annual Review of Applied Linguistics*, *41*, 118-125. https://doi.org/10.1017/S0267190521000052

Pearlmutter, N. J., Garnsey, S. M., & Bock, K. (1999). Agreement processes in sentence comprehension. *Journal of Memory and Language*, *41*(3), 427-456. https://doi.org/10.1006/jmla.1999.2653

Peirce, J., Gray, J. R., Simpson, S., MacAskill, M., Höchenberger, R., Sogo, H., Kastman, E., & Lindeløv, J. K. (2019). PsychoPy2: Experiments in behavior made easy. *Behavior Research Methods*, *51*(1), 195-203. https://doi.org/10.3758/s13428-018-01193-y

Perruchet, P. (2019). What mechanisms underlie implicit statistical learning? Transitional probabilities versus chunks in language learning. *Topics in Cognitive Science*, *11*(3), 520-535. https://doi.org/10.1111/tops.12403

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. https://doi.org/10.1177/001316447603600216

Pierce, L. J., Genesee, F., Delcenserie, A., & Morgan, G. (2017). Variations in phonological working memory: Linking early language experiences and language learning outcomes.

Applied Psycholinguistics, *38*(6), 1265-1300. https://doi.org/10.1017/S0142716417000236

Pimsleur, P. (1966). *Pimsleur Language Aptitude Battery*. Harcourt Brace Jovanovich.

Pinker, S., & Ullman, M. T. (2002). The past and future of the past tense. *Trends in Cognitive Sciences*, *6*(11), 456-463. https://doi.org/10.1016/S1364-6613(02)01990-3

Posit team. (2024). *RStudio: Integrated development environment for R*. In (Version 2024.04.1) Posit Software, PBC. http://www.posit.co/

Pranjković, I. (2000). Izražavanje neodređenosti/određenosti imenica u hrvatskome jeziku. *Riječki filološki dani*, *3*(2000), 343-350.

Quallo, M. M., Price, C. J., Ueno, K., Asamizuya, T., Cheng, K., Lemon, R. N., & Iriki, A. (2009). Gray and white matter changes associated with tool-use learning in macaque monkeys. *Proceedings of the National Academy of Sciences*, *106*(43), 18379-18384. https://doi.org/10.1073/pnas.0909751106

R Core Team. (2021). *R: A language and environment for statistical computing*. In (Version 2021.09.2) R Foundation for Statistical Computing. https://www.R-project.org/

R Core Team. (2024). R: A language and environment for statistical computing. In: R Foundation for Statistical Computing.

Reber, A. S. (1967). Implicit learning of artificial grammars. *Journal of Verbal Learning and Verbal Behavior*, *6*(6), 855-863. https://doi.org/10.1016/S0022-5371(67)80149-X

Reber, A. S. (1992). The cognitive unconscious: An evolutionary perspective. *Consciousness and Cognition*, *1*(2), 93-133. https://doi.org/10.1016/1053-8100(92)90051-B

Reber, A. S. (1996). *Implicit learning and tacit knowledge: An essay on the cognitive unconscious* (Vol. 19). Oxford University Press. https://doi.org/10.1093/acprof:oso/9780195106589.001.0001

Reber, A. S., & Allen, R. (2000). Individual differences in implicit learning: Implications for the evolution of consciousness. In *Individual differences in conscious experience.* (pp. 227-247). John Benjamins Publishing Company. https://doi.org/10.1075/aicr.20.11reb

Reber, A. S., Walkenfeld, F. F., & Hernstadt, R. (1991). Implicit and explicit learning: individual differences and IQ. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *17*(5), 888–896. https://doi.org/10.1037//0278-7393.17.5.888

Rebuschat, P. (2013). Measuring implicit and explicit knowledge in second language research. *Language Learning*, *63*(3), 595-626. https://doi.org/10.1111/lang.12010

Rebuschat, P. (2022). Implicit learning and language acquisition. In. Oxford University Press. https://doi.org/10.1093/oso/9780197501573.003.0006

Rebuschat, P., Hamrick, P., Riestenberg, K., Sachs, R., & Ziegler, N. (2015). Triangulating measures of awareness: A contribution to the debate on learning without awareness. *Studies in Second Language Acquisition*, *37*(2), 299-334. https://doi.org/10.1017/S0272263115000145

Reed, J., & Johnson, P. (1994). Assessing implicit learning with indirect tests: Determining what is learned about sequence structure. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *20*(3), 585-594. https://doi.org/10.1037/0278-7393.20.3.585

Révész, A. (2012). Working memory and the observed effectiveness of recasts on different L2 outcome measures. *Language Learning*, *62*(1), 93-132. https://doi.org/10.1111/j.1467-9922.2011.00690.x

Roberts, L., & Liszka, S. A. (2013). Processing tense/aspect-agreement violations on-line in the second language: A self-paced reading study with French and German L2 learners of English. *Second Language Research*, *29*(4), 413-439. https://doi.org/10.1177/0267658313503171

Robinson, P. (2002). Learning conditions, aptitude complexes, and SLA. In P. Robinson (Ed.), *Individual differences and instructed language learning* (pp. 113 – 133). John Benjamins Publishing Company. https://doi.org/10.1075/lllt.2.08rob

Robinson, P. (2005). Aptitude and second language acquisition. *Annual Review of Applied Linguistics*, *25*, 46-73. https://doi.org/10.1017/S0267190505000036

Robinson, P. (2007). Aptitudes, abilities, contexts, and practice. In R. M. DeKeyser (Ed.), *Practice in a second language: Perspectives from applied linguistics and cognitive psychology* (pp. 256-286). Cambridge University Press.

Robinson, P. (2012). Individual differences, aptitude complexes, SLA processes, and aptitude test development. In M. Pawlak (Ed.), *New perspectives on individual differences in language learning and teaching* (pp. 57-75). Springer.

Roehr, K. (2008). Metalinguistic knowledge and language ability in university-level L2 learners. *Applied Linguistics*, *29*(2), 173-199. https://doi.org/10.1093/applin/amm037

Roehr, K., & Gánem-Gutiérrez, G. A. (2009a). Metalinguistic knowledge: A stepping stone towards L2 proficiency In A. G. Benati (Ed.), *Issues in Second Language Proficiency* (pp. 79-94). Continuum.

Roehr, K., & Gánem-Gutiérrez, G. A. (2009b). The status of metalinguistic knowledge in instructed adult L2 learning. *Language Awareness*, *18*(2), 165-181. https://doi.org/10.1080/09658410902855854

Roehr-Brackin, K. (2015). Explicit knowledge about language in L2 learning: A usage-based perspective. In P. Rebuschat (Ed.), *Implicit and explicit learning of languages* (pp. 117-138). John Benjamins Publishing Company. https://doi.org/10.1075/sibil.48.06roe

Roehr-Brackin, K., Baranowska, K., Pavlekovic, R., & Scheffler, P. (2024). The role of individual learner differences in explicit language instruction. *The Modern Language Journal*, *108*(4), 815-845. https://doi.org/10.1111/modl.12963

Roehr-Brackin, K., Gánem-Gutiérrez, G. A., Olivera-Smith, L., & Torres-Marín, M. T. (2021). Are individual differences in cognitive abilities and stylistic preferences related to multilingual adults' performance in explicit learning conditions? *Language Awareness*, *30*(4), 391-412. https://doi.org/10.1080/09658416.2021.1969404

Roehr-Brackin, K., Loaiza, V., & Pavlekovic, R. (2023). Language learning aptitude in older adults. *Journal of the European Second Language Association*. https://doi.org/10.22599/jesla.93

Rogers, V., Meara, P., Barnett-Legh, T., Curry, C., & Davie, E. (2017). Examining the LLAMA aptitude tests. *Journal of the European Second Language Association*. https://doi.org/10.22599/jesla.24

Rogers, V., Meara, P., & Rogers, B. (2023). Testing language aptitude: LLAMA evolution and refinement. In Z. E. Wen, P. Skehan, & R. L. Sparks (Eds.), *Language Aptitude Theory and Practice* (pp. 47-72). Cambridge University Press. https://doi.org/10.1017/9781009076463.005

Rogers, V. E., Meara, P., Aspinall, R., Fallon, L., Goss, T., Keey, E., & Thomas, R. (2016). Testing aptitude: Investigating Meara's (2005) LLAMA tests. *EUROSLA Yearbook*, *16*(1), 179-210. https://doi.org/10.1075/eurosla.16.07rog

Rogosa, D. (1981). On the relationship between the Johnson-Neyman region of significance and statistical tests of parallel within-group regressions. *Educational and Psychological Measurement*, *41*(1), 73-84. https://doi.org/10.1177/001316448104100108

Russell, J., & Spada, N. (2006). *The effectiveness of corrective feedback for the acquisition of L2 grammar: A meta-analysis of the research*. In J. M. Norris & L. Ortega (Eds.), *Synthesizing research on language learning and teaching*. John Benjamins.

Saeed, J. (1997). *Semantics*. Blackwell.

Sáfár, A., & Kormos, J. (2008). Revisiting problems with foreign language aptitude. *International Review of Applied Linguistics in Language Teaching*, *46*(2), 113-136. https://doi.org/doi:10.1515/IRAL.2008.005

Sagarra, N. (2007). Working memory and L2 processing of redundant grammatical forms. In Z. Han (Ed.), *Understanding second language process* (pp. 133-147). Second Language Acquisition.

Sagarra, N. (2017). Longitudinal effects of working memory on L2 grammar and reading abilities. *Second Language Research*, *33*(3), 341-363. https://doi.org/10.1177/0267658317690577

Saint-Cyr, J. A., Taylor, A. E., & Lang, A. E. (1988). Procedural learning and neostyrial dysfunction in man. *Brain*, *111*(4), 941-960. https://doi.org/10.1093/brain/111.4.941

Saito, K., Sun, H. U. I., & Tierney, A. (2019). Explicit and implicit aptitude effects on second language speech learning: Scrutinizing segmental and suprasegmental sensitivity and performance via behavioural and neurophysiological measures. *Bilingualism: Language and Cognition*, *22*(5), 1123-1140. https://doi.org/10.1017/S1366728918000895

Sanz, C., Lin, H.-J., Lado, B., Stafford, C. A., & Bowden, H. W. (2016). One size fits all? Learning conditions and working memory capacity in ab initio language development. *Applied Linguistics*, *37*(5), 669-692. https://doi.org/10.1093/applin/amu058

Sasaki, M. (1996). *Second language proficiency, foreign language aptitude, and intelligence*. Lang.

Schloerke, B., Cook, D., Larmarange, J., Briatte, F., Marbach, M., Thoen, E., Elberg, A., & Crowley, J. (2024). *GGally: Extension to 'ggplot2'*. In https://CRAN.R-project.org/package=GGally

Schmidt, R. (1990). The role of consciousness in second language learning. *Applied Linguistics*, *11*(2), 129-158. https://doi.org/10.1093/applin/11.2.129

Schmidt, R. (2001). Attention. In P. Robinson (Ed.), *Cognition and second language instruction* (pp. 3-32). Cambridge University Press.

Schmidt, R. (2007). Attention. In P. Robinson (Ed.), *Cognition and second language instruction* (3rd ed., pp. 3-32). Cambridge University Press.

Schmidt, R. (2012). *Attention, awareness, and individual differences in language learning*. In W. Chan, K. Chin, S. Bhatt, & I. Walker (Eds.), *Perspectives on individual characteristics and foreign language education* (1st ed.). De Gruyter Mouton. https://doi.org/10.1515/9781614510932

Schmidt, R., & Frota, S. (1986). Developing basic conversational ability in a second language: A case-study of an adult learner. In R. R. Day (Ed.), *Talking to learn: Conversation in second language acquisition* (pp. 237-326). Newbury House.

Serafini, E. J., & Sanz, C. (2016). Evidence for the decreasing impact of cognitive ability on second language development as proficiency increases. *Studies in Second Language Acquisition*, *38*(4), 607-646. https://doi.org/10.1017/S0272263115000327

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

Service, E., & Kohonen, V. (1995). Is the relation between phonological memory and foreign language learning accounted for by vocabulary acquisition? *Applied Psycholinguistics*, *16*(2), 155-172. https://doi.org/10.1017/S0142716400007062

Shah, P., & Miyake, A. (1996). The separability of working memory resources for spatial thinking and language processing: An individual differences approach. *Journal of Experimental Psychology: General*, *125*(1), 4-27. https://doi.org/10.1037/0096-3445.125.1.4

Shallice , T. (1982). Specific impairments of planning. *Phil. Trans. R. Soc. Lond. B*, *298*, 199-209. https://doi.org/10.1098/rstb.1982.0082

Sharwood Smith, M. (1981). Consciousness-raising and the second language learner. *Applied Linguistics*, *II*(2), 159-168. https://doi.org/10.1093/applin/II.2.159

Siegelman, N., Bogaerts, L., Elazar, A., Arciuli, J., & Frost, R. (2018). Linguistic entrenchment: Prior knowledge impacts statistical learning performance. *Cognition*, *177*, 198-213. https://doi.org/10.1016/j.cognition.2018.04.011

Simon, H. A. (1975). The functional equivalence of problem solving skills. *Cognitive Psychology*, *7*(2), 268-288. https://doi.org/10.1016/0010-0285(75)90012-2

Skehan, P. (1998). *A cognitive approach to language learning*. Oxford University Press.

Skehan, P. (2002). Theorising and updating aptitude. In P. Robinson (Ed.), *Individual differences and instructed language learning* (pp. 69-93). John Benjamins Publishing Company. https://doi.org/10.1075/lllt.2.06ske

Skehan, P. (2012). Language aptitude. In S. M. Gass & A. Mackey (Eds.), *The Routledge handbook of second language acquisition*. *Routledge, Taylor & Francis Group*.

Skehan, P. (2016). Foreign language aptitude, acquisitional sequences, and psycholinguistic processes. In G. Granena, D. O. Jackson, & Y. Yilmaz (Eds.), *Cognitive individual differences in second language processing and acquisition*. John Benjamins Publishing Company.

Skehan, P. (2019). Language aptitude implicates language and cognitive skills. In Z. Wen, P. Skehan, A. Biedrón, S. Li, & R. L. Sparks (Eds.), *Language aptitude: Advancing theory, testing, research and practice*. Routledge, Taylor & Francis Group.

Snow, R. E. (1987). Aptitude complexes. In R. Snow, E. & M. J. Farr (Eds.), *Aptitude, learning and instruction* (Vol. 3: Conative and affective process analysis, pp. 11-34). Erlbaum.

Snow, R. E. (1994). Abilities in academic tasks. In R. J. Stenberg & R. K. Wagner (Eds.), *Mind in context: Interactionist perspectives on human intelligence* (pp. 3-37). Cambridge University Press.

Sorace, A. (1985). Metalinguistic knowledge and language use in acquisition-poor environments*. *Applied Linguistics*, *6*(3), 239-254. https://doi.org/10.1093/applin/6.3.239

Sorace, A. (1996). The use of acceptability judgments in second language acquisition research. In W. C. Ritchie & T. K. Bhatia (Eds.), *Handbook of second language acquisition* (pp. 375-409). Academic Press.

Spada, N. (1997). Form-focussed instruction and second language acquisition: A review of classroom and laboratory research. *Language Teaching*, *30*(2), 73-87. https://doi.org/10.1017/S0261444800012799

Spada, N., & Tomita, Y. (2010). Interactions between type of instruction and type of language feature: A meta-analysis. *Language Learning*, *60*(2), 263-308. https://doi.org/10.1111/j.1467-9922.2010.00562.x

Sparks, R. L., & Ganschow, L. (1991). Foreign language learning differences: Affective or native language aptitude differences? *The Modern Language Journal*, *75*(1), 3-16. https://doi.org/10.1111/j.1540-4781.1991.tb01076.x

Sparks, R. L., & Ganschow, L. (2001). Aptitude for learning a foreign language. *Annual Review of Applied Linguistics*, *21*, 90-111. https://doi.org/10.1017/S026719050100006X

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. https://doi.org/10.1111/j.1540-4781.2011.01176.x

Speciale, G., Ellis, N. C., & Bywater, T. (2004). Phonological sequence learning and short-term store capacity determine second language vocabulary acquisition. *Applied Psycholinguistics*, *25*(2), 293-321. https://doi.org/10.1017/S0142716404001146

Sprenger, A. M., Atkins, S. M., Bolger, D. J., Harbison, J. I., Novick, J. M., Chrabaszcz, J. S., Weems, S. A., Smith, V., Bobb, S., Bunting, M. F., & Dougherty, M. R. (2013). Training working memory: Limits of transfer. *Intelligence*, *41*(5), 638-663. https://doi.org/10.1016/j.intell.2013.07.013

Stanovich, K. E. (2009). Distinguishing the reflective, algorithmic, and autonomous minds: Is it time for a tri-process theory? In J. Evans & K. Frankish (Eds.), *In two minds : Dual processes and beyond*. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199230167.003.0003

Stanovich, K. E., & West, R. F. (2000). Individual differences in reasoning: Implications for the rationality debate? *Behavioral and Brain Sciences*, *23*(5), 645-665. https://doi.org/10.1017/S0140525X00003435

Sternberg, R. J. (1997). *Successful intelligence: How practical and creative intelligence determine success in life*. Plume.

Sternberg, R. J. (2002). The theory of successful intelligence and its implications for language aptitude testing In P. Robinson (Ed.), *Individual differences and instructed language learning* (pp. 13-43). John Benjamins Publishing Company.

Sternberg, R. J., & Grigorenko, E. L. (2002). The theory of successful intelligence as a basis for gifted education. *Gifted Child Quarterly*, *46*(4), 265-277. https://doi.org/10.1177/001698620204600403

Suzuki, Y. (2017). Validity of new measures of implicit knowledge: Distinguishing implicit knowledge from automatized explicit knowledge. *Applied Psycholinguistics*, *38*(5), 1229-1261. https://doi.org/10.1017/S014271641700011X

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. https://doi.org/10.1017/S0272263120000704

Suzuki, Y., & DeKeyser, R. M. (2015). Comparing elicited imitation and word monitoring as measures of implicit knowledge. *Language Learning*, *65*(4), 860-895. https://doi.org/10.1111/lang.12138

Suzuki, Y., & DeKeyser, R. M. (2017). The interface of explicit and implicit knowledge in a second language: Insights from individual differences in cognitive aptitudes. *Language Learning*, *67*(4), 747-790. https://doi.org/10.1111/lang.12241

Suzuki, Y., Jeong, H., Cui, H., Okamoto, K., Kawashima, R., & Sugiura, M. (2023). fMRI reveals the dynamic interface between explicit and implicit knowledge recruited during elicited imitation task. *Research Methods in Applied Linguistics*, *2*(2), 100051. https://doi.org/10.1016/j.rmal.2023.100051

Tabachnick, B. G., & Fidell, L. S. (2014). *Using multivariate statistics* (6th ed., Pearson new international ed. ed.). Pearson.

Tagarelli, K. M., Shattuck, K. F., Turkeltaub, P. E., & Ullman, M. T. (2019). Language learning in the adult brain: A neuroanatomical meta-analysis of lexical and grammatical learning. *NeuroImage*, *193*, 178-200. https://doi.org/10.1016/j.neuroimage.2019.02.061

Tarone, E. (1988). *Variation in interlanguage*. Edward Arnold.

Taubert, M., Draganski, B., Anwander, A., Müller, K., Horstmann, A., Villringer, A., & Ragert, P. (2010). Dynamic properties of human brain structure: Learning-related changes in cortical areas and associated fiber connections. *The Journal of Neuroscience*, *30*(35), 11670. https://doi.org/10.1523/JNEUROSCI.2567-10.2010

Teimouri, Y., Goetze, J., & Plonsky, L. (2019). Second language anxiety and achievement: A meta-analysis. *Studies in Second Language Acquisition*, *41*(2), 363-387. https://doi.org/10.1017/S0272263118000311

Tomak, I. (2019). *Contributions of type of instruction, individual differences in cognitive ability and age to the development of explicit and implicit knowledge of L2 English articles: A study with Russian learners of English*. Unpublished PhD thesis, University of Essex.

Trahan, D. E., & Larrabee, G. J. (1988). Continuous visual memory test. In. Odessa, FL: Assessment Resources.

Turner, M. L., & Engle, R. W. (1989). Is working memory capacity task dependent? *Journal of Memory and Language*, *28*(2), 127-154. https://doi.org/10.1016/0749-596X(89)90040-5

Tyler, M. D. (2001). Resource consumption as a function of topic knowledge in nonnative and native comprehension. *Language Learning*, *51*(2), 257-280. https://doi.org/10.1111/1467-9922.00155

Ullman, M. T. (2004). Contributions of memory circuits to language: The declarative/procedural model. *Cognition*, *92*(1-2), 231–270. https://doi.org/10.1016/j.cognition.2003.10.008

Ullman, M. T. (2016). The declarative/procedural model: A neurobiological model of language learning, knowledge, and use. In G. Hickok & S. L. Small (Eds.), *Neurobiology of Language* (pp. 953-968). Academic Press. https://doi.org/10.1016/B978-0-12-407794-2.00076-6

Ullman, M. T. (2020). The declarative/procedural model: A neurobiologically motivated theory of first and second language. In B. VanPatten, G. D. Keating, & S. Wulff (Eds.), *Theories in second language acquisition: An introduction* (3rd ed., pp. 128-161). Routledge, Taylor & Francis Group. https://doi.org/10.4324/9780429503986-7

Unsworth, N., Heitz, R. P., Schrock, J. C., & Engle, R. W. (2005). An automated version of the operation span task. *Behavior Research Methods*, *37*(3), 498-505. https://doi.org/10.3758/BF03192720

Unterrainer, J. M., Rahm, B., Leonhart, R., Ruff, C. C., & Halsband, U. (2003). The Tower of London: The impact of instructions, cueing, and learning on planning abilities. *Cognitive Brain Research*, *17*(3), 675-683. https://doi.org/10.1016/S0926-6410(03)00191-5

Vafaee, P., Suzuki, Y., & Kachisnke, I. (2017). Validating grammaticality judgment tests: Evidence from two new psycholinguistic measures. *Studies in Second Language Acquisition*, *39*(1), 59-95. https://doi.org/10.1017/S0272263115000455

van Buuren, S., & Groothuis-Oudshoorn, K. (2011). mice: Multivariate imputation by chained equations in R. *Journal of Statistical Software*, *45*(3), 1 - 67. https://doi.org/10.18637/jss.v045.i03

van den Noort, M. W. M. L., Bosch, P., & Hugdahl, K. (2006). Foreign language proficiency and working memory capacity. *European Psychologist*, *11*(4), 289-296. https://doi.org/10.1027/1016-9040.11.4.289

Walter, C. (2004). Transfer of reading comprehension skills to L2 is linked to mental representations of text and to L2 working memory. *Applied Linguistics*, *25*(3), 315-339. https://doi.org/10.1093/applin/25.3.315

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. https://doi.org/10.1080/713755607

Wen, Z. (2016). *Working memory and second language learning: Towards an integrated approach*. Multilingual matters. https://doi.org/10.21832/9781783095735

Wen, Z. (2019). Working memory as language aptitude. In Z. Wen, P. Skehan, A. Biedrón, S. Li, & R. L. Sparks (Eds.), *Language aptitude: Advancing theory, testing, research and practice*. Routledge, Taylor & Francis Group.

Wen, Z., Biedroń, A., & Skehan, P. (2017). Foreign language aptitude theory: Yesterday, today and tomorrow. *Language Teaching*, *50*(1), 1-31. https://doi.org/10.1017/S0261444816000276

Williams, J. N. (2005). Learning without awareness. *Studies in Second Language Acquisition*, *27*(2), 269-304. https://doi.org/10.1017/S0272263105050138

Williams, J. N. (2009). Implicit learning in second language acquisition. In W. C. Ritchie & T. K. Bhatia (Eds.), *The new handbook of second language acquisition* (2nd ed. ed.). Emerald.

Williams, J. N., & Lovatt, P. (2003). Phonological memory and rule learning. *Language Learning*, *53*(1), 67-121. https://doi.org/10.1111/1467-9922.00211

Woltz, D. J. (2003). Implicit cognitive processes as aptitudes for learning. *Educational Psychologist*, *38*(2), 95-104. https://doi.org/10.1207/S15326985EP3802_4

Yalçın, Ş., & Spada, N. (2016). Language aptitude and grammatical difficulty: An EFL classroom-based study. *Studies in Second Language Acquisition*, *38*(2), 239-263. https://doi.org/10.1017/S0272263115000509

Yi, W. (2018). Statistical sensitivity, cognitive aptitudes, and processing of collocations. *Studies in Second Language Acquisition*, *40*(4), 831-856. https://doi.org/10.1017/S0272263118000141

Yilmaz, Y., & Granena, G. (2019). Cognitive individual differences as predictors of improvement and awareness under implicit and explicit feedback conditions. *The Modern Language Journal*, *103*(3), 686-702. https://doi.org/10.1111/modl.12587

Yoshimura, Y. (2001). The role of working memory in language aptitude. In X. Bonch-Bruevich, W. J. Crawford, J. Hellermann, C. Higgins, & H. Nguyen (Eds.), *The Past, Present, and Future of Second Language Research*. Cascadilla Press.

Zergollern-Miletić, L. (2008). *Kategorija određenosti i neodređenosti u engleskom i hrvatskom jeziku*. University of Zagreb.

Zhang, R. (2015). Measuring university-level L2 learners' implicit and explicit linguistic knowledge. *Studies in Second Language Acquisition*, *37*(3), 457-486. https://doi.org/10.1017/S0272263114000370

Appendix 1

Table A: Self-paced reading items

Item	List 1	List 2
	Articles	
1	They wanted to open a **café** where they could sell good coffee.	They wanted to open **café** where they could sell good coffee.
2	Roger works for a **firm** in Midtown Manhattan.	Roger works for **firm** in Midtown Manhattan.
3	Can you tell me if there is an **underground** station near here?	Can you tell me if there is **underground** station near here?
4	There were **cats** in every room.	There were a **cats** in every room.
5	In Australia, **farmers** are the biggest users of groundwater.	In Australia, a **farmers** are the biggest users of groundwater.
6	Her coat is made of pure **wool** from New Zealand.	Her coat is made of a pure **wool** from New Zealand.
7	In the 9th century, the **capital** of England was Winchester.	In the 9th century, **capital** of England was Winchester.
8	Scientists say that the **sun** is the most important source of energy for life on Earth.	Scientists say that a **sun** is the most important source of energy for life on Earth.
9	Greece is located in the **Mediterranean** in Southern Europe.	Greece is located in **Mediterranean** in Southern Europe.
10	It is said that the **Amazon** river is one vast highway.	It is said that **Amazon** river is one vast highway.
11	Recent reports claim that the **tiger** is in danger of becoming extinct.	Recent reports claim that **tiger** is in danger of becoming extinct.
12	My sister has been playing the **violin** for three years.	My sister has been playing a **violin** for three years.
13	Carol was wearing beautiful **bracelet** that summer evening.	Carol was wearing a beautiful **bracelet** that summer evening.
14	Usually, swimming **pool** lies within the ground.	Usually, a swimming **pool** lies within the ground.
15	We saw **bear** while driving yesterday.	We saw a **bear** while driving yesterday.
16	It is known that a **nurses** work very hard.	It is known that **nurses** work very hard.
17	I read somewhere that a **metals** are mostly shiny.	I read somewhere that **metals** are mostly shiny.
18	Their kitchen always smelled of the **rice** and curry.	Their kitchen always smelled of **rice** and curry.
19	People used to think an **earth** was flat.	People used to think the **earth** was flat.
20	Leonardo da Vinci painted a *Mona* ***Lisa*** in the 16th century.	Leonardo da Vinci painted the *Mona* ***Lisa*** in the 16th century.
21	His father liked skiing in an **Alps** and diving in California.	His father liked skiing in the **Alps** and diving in California.

22	Uniquely in **Bahamas**, all species of sharks are protected.	Uniquely in the **Bahamas**, all species of sharks are protected.
23	Their analysis confirmed that an **elephant** is the largest land animal on Earth.	Their analysis confirmed that the **elephant** is the largest land animal on Earth.
24	Some historic reports show **wheel** was invented at least two dozen times.	Some historic reports show the **wheel** was invented at least two dozen times.
	Passive	
25	How is this word **used** in a sentence?	How is this word **use** in a sentence?
26	In Italy, a large number of people is **employed** by the government.	In Italy, a large number of people is **employ** by the government.
27	In the next step, DNA samples are **taken** from each victim.	In the next step, DNA samples are **taking** from each victim.
28	The entire camp was **woken** up by a loud noise during the night.	The entire camp was **wake** up by a loud noise during the night.
29	The roof of the building was **damaged** in a storm a few days ago.	The roof of the building was **damage** in a storm a few days ago.
30	This house was **built** by my grandfather.	This house was **building** by my grandfather.
31	A lot of money was **stolen** in the robbery.	A lot of money was **stealing** in the robbery.
32	The date of the meeting has been **changed** and everyone was notified soon after.	The date of the meeting has been **change** and everyone was notified soon after.
33	The windows are very dirty since they haven’t been **cleaned** for ages.	The windows are very dirty since they haven’t been **cleaning** for ages.
34	I didn’t know that our conversation was being **recorded** this entire time.	I didn’t know that our conversation was being **recording** this entire time.
35	Kayaks can be **rented** at various shops around the island.	Kayaks can be **rent** at various shops around the island.
36	All Beatles records can be **borrowed** from the central library.	All Beatles records can be **borrow** from the central library.
37	About a third of world’s land surface is **cover** with forests.	About a third of world’s land surface is **covered** with forests.
38	The garbage is **collect** every day.	The garbage is **collected** every day.
39	With a new system, any problems are **reporting** within an hour.	With a new system, any problems are **reported** within an hour.
40	The movie ET was **direct** by Steven Spielberg.	The movie ET was **directed** by Steven Spielberg.
41	There was a meeting yesterday and I wasn’t **tell** anything about it.	There was a meeting yesterday and I wasn’t **told** anything about it.
42	The novel *Anna Karenina* was **writing** by Leo Tolstoy.	The novel *Anna Karenina* was **written** by Leo Tolstoy.
43	The thieves were **arresting** in the early afternoon.	The thieves were **arrested** in the early afternoon.
44	The decision has been **make** and the company is taking a different direction.	The decision has been **made** and the company is taking a different direction.
45	I couldn’t believe my hamster has been **letting** out from the cage.	I couldn’t believe my hamster has been **let** out from the cage.

46	Tom was being **questioning** at the police station when I called him.	Tom was being **questioned** at the police station when I called him.
47	Supplemental information can be **find** on their website.	Supplemental information can be **found** on their website.
48	Additional questions can be **ask** on the forum.	Additional questions can be **asked** on the forum.
	Past	
49	When my brother was younger, he didn't **like** fish and chips.	When my brother was younger, he didn't **liked** fish and chips.
50	Kate received her paycheck yesterday but didn't **spend** one penny.	Kate received her paycheck yesterday but didn't **spent** one penny.
51	When my cousin was a child, she didn't **want** any pets.	When my cousin was a child, she didn't **wanted** any pets.
52	Did it **rain** on Sunday morning?	Did it **rained** on Sunday morning?
53	When I was a child, I **visited** my grandma every weekend.	When I was a child, I **visit** my grandma every weekend.
54	My father **applied** for this job four times before they called him back.	My father **apply** for this job four times before they called him back.
55	After he crossed the road, he **pulled** a gun and entered the big bank.	After he crossed the road, he **pulling** a gun and entered the big bank.
56	They were in the middle of the movie when someone **knocked** on the door.	They were in the middle of the movie when someone **knocking** on the door.
57	Most of my childhood I **spent** in a small Scottish city.	Most of my childhood I was **spent** in a small Scottish city.
58	When he was younger, Dwight **had** two black cats.	When he was younger, Dwight was **had** two black cats.
59	Each night we **explored** a little more of the area.	Each night we were **explored** a little more of the area.
60	He heard the news, grabbed his jacket and **ran** into the storm.	He heard the news, grabbed his jacket and was **ran** into the storm.
61	After the movie, we were hungry, but we didn't **went** to a restaurant.	After the movie, we were hungry, but we didn't **go** to a restaurant.
62	What did you **did** on the weekend?	What did you **do** on the weekend?
63	Laura received her paycheck, went to a travel agency, but didn't **bought** a plane ticket.	Laura received her paycheck, went to a travel agency, but didn't **buy** a plane ticket.
64	Kelly ordered her food but didn't **asked** where the bathroom was.	Kelly ordered her food but didn't **ask** where the bathroom was.
65	Regularly every summer, Janet **fall** in love with a different man.	Regularly every summer, Janet **fell** in love with a different man.
66	Jane was washing the dishes when her husband **call** to check on her.	Jane was washing the dishes when her husband **called** to check on her.
67	Yesterday, I **playing** a tennis match with my friend.	Yesterday, I **played** a tennis match with my friend.
68	Last night, the police **stopping** me on my way home.	Last night, the police **stopped** me on my way home.

69	While I was looking for my car keys, I was **stumbled** upon an old photo.	While I was looking for my car keys, I **stumbled** upon an old photo.
70	We were watching the movie when her dog was **jumped** out of the window.	We were watching the movie when her dog **jumped** out of the window.
71	Every summer, my family was **rented** a villa in France.	My family **rented** a villa in France every summer.
72	After they had had an argument, she was packed **her** things and left.	After they had had an argument, she packed **her** things and left.
	Fillers	
73	Above all, you must help each other.	Above all, you must help each other.
74	She loves him now more than she did before.	She loves him now more than she did before.
75	For some reason, I am wide awake and can't fall asleep.	For some reason, I am wide awake and can't fall asleep.
76	She is not afraid of anything.	She is not afraid of anything.
77	Mary seems to be bored with the game.	Mary seems to be bored with the game.
78	He stuck with his own theory.	He stuck with his own theory.
79	I never imagined anything like this.	I never imagined anything like this.
80	I remember what he said.	I remember what he said.
81	He really wants to buy a new motorcycle.	He really wants to buy a new motorcycle.
82	Algebra is a branch of mathematics.	Algebra is a branch of mathematics.
83	I think it's time for me to contact her.	I think it's time for me to contact her.
84	If there is anything I can do for you, please let me know.	If there is anything I can do for you, please let me know.
85	She walked as fast as she could to catch up with him.	She walked as fast as she could to catch up with him.
86	She advised him to use a bicycle.	She advised him to use a bicycle.
87	The virus had powers none of us knew existed.	The virus had powers none of us knew existed.
88	For your own safety, never ride in a car with a drunk driver.	For your own safety, never ride in a car with a drunk driver.
89	Trains are running on schedule.	Trains are running on schedule.
90	Peanuts don't grow on trees, but cashews do.	Peanuts don't grow on trees, but cashews do.
91	He put a piece of gauze on the wound to stop the bleeding.	He put a piece of gauze on the wound to stop the bleeding.
92	There are few things better in life than a slice of pie.	There are few things better in life than a slice of pie.
93	The secret code they created made no sense, even to them.	The secret code they created made no sense, even to them.
94	Japanese knives stay incredibly sharp and precise.	Japanese knives stay incredibly sharp and precise.
95	The waves were crashing on the shore; it was a lovely sight.	The waves were crashing on the shore; it was a lovely sight.
96	In the end, they sold their house for more than they had hoped for.	In the end, they sold their house for more than they had hoped for.

Table B: Elicited imitation items

Item	Sentence
	Articles
1	A woman with her hands full says to her colleague: “Open the door, please, would you?
2	I haven’t been to the cinema this year.
3	I read a book about New York. The author, however, was from Arizona.
4	We rented a boat last summer. Unfortunately, the boat hit another boat and sank.
5	In July and August the sun shines for longer than in November and December.
6	Countries near to the equator receive more sunlight all year round.
7	Contrary to popular belief, stealing is not something the homeless usually do.
8	That house had bugs down in the basement.
9	Mary is not tall but she plays basketball very well.
10	John’s wife died of cancer back in 1996.
11	My neighbor has a Spanish name, but in fact she’s English.
12	My desktop computer has a screen that is twice as big as my laptop’s.
13	The wife asked the husband if he could pass her salt because she couldn’t reach it.
14	The manager asks her secretary, “Could you please update schedule later today?”
15	Fred bought a car on Monday. On Wednesday, he crashed car while going to work.
16	We went to a wedding. Bride wore a long veil and the groom a black tuxedo.
17	Einstein thought universe was standing still.
18	St. Petersburg used to be capital before the Russian revolution.
19	Like everyone else, rich spend more on housing when they have more money.
20	It’s annoying when a people throw rubbish on the ground.
21	Someday, I think that a computers will replace people.
22	I am going to the bed really soon.
23	I am going away for week or two next month.
24	My sister is doctor who works at the local hospital.
	Passive
25	The water in the swimming pool is refreshed every week in that hotel.
26	Refrigerated foods should always be kept between 1 °C and 5 °C.
27	Facebook is used by many young people.
28	Stamps can be purchased online or at the post office.
29	During the meeting, major concerns were raised about the future of the company.
30	All flights were cancelled because of Coronavirus.
31	While I was on holiday, my camera was stolen from the hotel room.
32	The winning goal was scored from an offside position.
33	The road has been fixed twice this year.
34	That window has been broken three times this month.
35	A new bridge is being built across the river.
36	The decision will not be made until the next meeting.

37	Most electronic goods are make either in China or Japan.
38	Family names are often change after marriage in Southern Europe.
39	Lithium is using worldwide for treatment of bipolar disorder.
40	Most of the Himalayan peaks are covering completely in snow.
41	My room was repaint last year during the summer.
42	The car was repair soon after the accident.
43	The pyramids were building over 2000 years ago.
44	Several trees were planting here last summer.
45	In school canteens, fast food has been forbid since April 2010.
46	Sara has been promote four times in her career.
47	The old church is being demolishing because of the new law.
48	The situation is serious, and something has to be doing before it's too late.
	Past
49	I knew Dr. Smith was busy, but I didn't want anyone else looking at the wound.
50	We went to the café and had a drink, but we didn't eat anything sweet.
51	The film wasn't very good. I didn't enjoy watching it.
52	He didn't like talking to the police yesterday.
53	I lived in Rome for two years. Then I left Italy and moved to Japan.
54	I didn't try her cake, but I had pancakes instead.
55	We couldn't afford paying the bills, so we sold the house.
56	The police checked the house, but they found nothing illegal inside.
57	I rang the doorbell and a woman opened completely shocked.
58	I tried to understand the movie, but the actors spoke extremely fast.
59	Yesterday, I washed several new shirts.
60	The window was open and a bird flew inside Jack's room.
61	The man said something to the woman, but she didn't heard what he said.
62	Jim asked me a question, but I didn't knew how to answer.
63	The bed was very uncomfortable, so I didn't slept well yesterday.
64	Yesterday, I started working in the morning and didn't finished until after midnight.
65	I didn't like the hotel since the room isn't very clean.
66	I learned how to drive because my dad teaches me some years ago.
67	When man invented the wheel, life changing dramatically for everyone.
68	We were tired of this place, so we packed and leaving everyone we knew.
69	Mark stopped smoking last month, and he was started playing tennis again.
70	Mozart was wrote more than 600 pieces of music.
71	Dave fell down the stairs and was injured himself badly as a consequence.
72	Everyone was celebrated when the war ended.

Table C: Gap-fill items

Item	Sentence	Correct	Distractor 1	Distractor 2
		Articles		
1	The highest mountain in ______ Andes is Aconcagua.	the	an	no article
2	Helen was listening to ______ music when I arrived.	no article	a	the
3	We went to ______ very nice restaurant last weekend.	a	the	no article
4	Buenos Aires is ______ capital of Argentina.	the	a	no article
5	My brother works for ______ insurance company.	an	the	no article
6	Can you tell me if there is ______ bank near here?	a	the	no article
7	Do you know anybody who collects ______ stamps?	no article	a	the
8	Lisa doesn't usually wear ______ jewelry.	no article	a	the
9	Jane was wearing ______ beautiful ring.	a	the	no article
10	Does this city have ______ airport?	an	the	no article
11	I like ______ football. It's a good game.	no article	a	the
12	The heart pumps ______ blood through the body.	no article	a	the
13	I learnt to play ______ piano when I was a child, but gave it up during college.	the	a	no article
14	______ Moon is Earth's only natural satellite.	The	A	no article
15	I am going away at ______ end of this month.	the	an	no article
16	Queen Elizabeth II is ______ Queen of England.	the	a	no article
17	The President rules the country from ______ White House.	the	a	no article
18	You will hardly ever see ______ polar bear outside the Arctic Circle.	the	a	no article
19	Can you play ______ guitar?	the	a	no article
20	______ Irish are voting on the new environmental policy tomorrow.	The	An	no article
21	When was ______ camera invented?	the	a	no article
22	______ Netherlands is a country known for tulip fields and windmills.	The	A	no article
23	______ Himalayas are spread across five countries.	The	A	no article
24	______ United States is the world's largest importer.	The	A	no article
25	______ giraffe is the tallest of all animals.	The	A	no article
		Passive		
26	Rio de Janeiro ______ as the most beautiful city in the Americas.	has been described	has being described	has described

27	A mysterious box ______ on the front porch yesterday.	was left	leave	leaves
28	There were some problems at first, but they seem to ______.	have been solved	solved	have solved
29	*Othello* ______ by Shakespeare.	was written	was write	wrote
30	The car was three years old, but it ______ very much.	hadn't been used	haven't been used	hadn't used
31	Your application ______ by the manager.	will be assessed	assesses	will assess
32	These yoga classes ______ by Susan since 2014.	have been taught	are taught	have taught
33	His company ______ worth almost three billion dollars.	is thought to be	is thinking	thinks to be
34	Cheese ______ from milk.	is made	is being made	make
35	In the US, presidential elections ______ every four years.	are held	are hold	hold
36	This plant is very rare. It ______ in very few places.	is found	found	finds
37	I never received the letter from you. It ______ to the wrong address.	was sent	was send	sent
38	I see the washing-up ______ again!	hasn't been done	wasn't done	hasn't done
39	The parcel ______ at midday.	is being delivered	delivers	is delivering
40	______ the reports ______ yet?	Have, been typed	Are, type	Have, typed
41	The building ______ date from the 13th century.	is believed to	is believing to	believes to
42	If you hadn't shouted at the policeman, you wouldn't ______.	have been arrested	arrest	have arrested
43	The American team ______ to win the race.	is expected	is expect	expect
44	A new school is being built. The old one is going to ______.	be knocked down	knocking down	knock down
45	Will our application for planning permission ______ ?	be granted	be grant	grant
46	I think that the old house ______.	has been demolished	demolished	has demolished
47	The election is next Sunday, and the full results will ______ on Tuesday.	be known	been known	know
48	There's somebody walking behind us. I think we ______.	are being followed	are being follow	are following
49	When I last visited, some new houses ______.	were being built	building	were building
50	My bicycle ______ .	has been stolen	stole	has stolen

Past

51	______ you ______ the film last night?	Did, see	Did, saw	Have, seen
52	Jerry ran to the car, ______ and raced off into the night.	jumped in	jump in	has jumped in
53	When Kate finished the race, she ______ exhausted.	felt	feel	has felt
54	Seventy cars were crossing the bridge when the supports ______ into the river.	collapsed	were collapsing	have collapsed
55	My grandfather ______ his paycheck on a weekly basis during the economic boom in the 1920s.	received	receive	has received
56	We ______ the lecture last week.	didn't understand	didn't understood	haven't understood
57	When ______ you ______ your last exam?	did, take	did, took	have, taken
58	When Jack arrived, he ______ us the news.	told	tells	has told
59	We were having dinner when Jane ______.	arrived	arrives	has arrived
60	I ______ to New York every month for three years when I was in my 20s.	travelled	travel	have travelled
61	______ Michelangelo ______ the *Mona Lisa*?	Did, paint	Did, painted	Has, painted
62	The bathroom ______ very strange last time I was there.	smelled	smell	has smelled
63	Henry ______ sad when I saw him this morning.	was	is	has been
64	My brother ______ for a visa six times before he got one.	applied	apply	has applied
65	He was trying to find the source of the Nile when he ______ in 1873.	died	was dying	has died
66	We lived in London for several years, but we ______ it much.	didn't like	didn't liked	haven't liked
67	Caroline ______ her family every day when she was on holiday.	phoned	phones	has phoned
68	MP3 players ______ when I was a child.	didn't exist	didn't existed	haven't existed
69	While I was driving to work this morning, I ______ a bear on the road.	saw	was seeing	have seen
70	They ______ the same test eight times before they found something.	ran	were running	have run
71	Pablo Picasso ______ during WW2.	lived	live	has lived
72	After Rory left home that morning, he ______ to work as usual.	went	goes	has gone
73	During the 2000s, the economic crisis ______ for two years.	lasted	lasts	has lasted
74	I was walking along the street when I suddenly ______ something behind me.	heard	was hearing	have heard
75	We arrived in Paris at noon and looked around the shops, but we ______ anything.	didn't buy	didn't bought	haven't bought

Appendix 2

Table A: Descriptive statistics for the LLAMA subtests

Subtest	n	k	M Raw	M Cor.	95% CI	SD Raw	SD Cor.	Skew	S-W (p)	α
LLAMA B	71	20	11.44	57.18	51.90 – 62.47	4.47	22.34	-0.50	0.017	0.81
LLAMA D	71	40	15.01	37.54	32.07 – 43.00	9.24	23.10	-0.02	0.022	0.72
LLAMA E	82	20	18.05	45.12	38.55 – 51.69	11.96	29.91	0.40	0.001	0.97
LLAMA F	82	20	66.83	50.63	45.75 – 55.51	29.32	22.22	-0.17	0.023	0.95

Note. n – number of participants; k – number of items

Table B: Descriptive statistics for the SRT task

Block	0	1	2	3	4	5	6	7	8
	Practice	Training (probable) condition							
M	475	459	456	449	448	441	428	427	426
SD	73	66	73	72	68	65	64	63	59
		Control (improbable) condition							
M		465	478	445	467	458	452	426	448
SD		68	76	68	72	65	65	61	64
		RT difference							
M		6.02	21.55	-4.09	18.96	16.86	23.67	-0.91	22.28
SD		27.50	26.10	26.10	28.13	27.34	29.65	23.37	25.01
S-W (p)		0.506	0.335	0.473	0.718	0.886	0.457	0.792	0.461

Table C: Paired-samples t-tests comparing RTs between training and control conditions by block

		M	SD	t	df	Sig.	d
B1 - (15)	B1 - (85)	7.70	30.21	2.324	82	0.023*	0.18
B2 - (15)	B2 - (85)	20.93	26.75	7.129	82	0.0001**	0.55
B3 - (15)	B3 - (85)	-4.46	26.57	-1.528	82	0.130	0.12[1]
B4 - (15)	B4 - (85)	18.46	28.27	5.951	82	0.0001**	0.46
B5 - (15)	B5 - (85)	15.95	27.64	5.223	81	0.0001**	0.41
B6 - (15)	B6 - (85)	23.98	29.30	7.457	82	0.0001**	0.57
B7 - (15)	B7 - (85)	-0.98	23.66	-0.377	81	0.707	0.03[1]
B8 - (15)	B8 - (85)	22.00	24.98	7.927	80	0.0001**	0.62
B4-8 (15)	B4-8 (85)	15.89	14.98	9.545	80	0.0001**	0.75

[1] In these blocks, the mean difference between conditions is in the “wrong” direction, i.e. participants are faster in the control than training condition

Table D: Factor loadings for a three-component solution (principal component analysis)

	Components		
Measure	1	2	3
LLAMA E	0.822	0.085	-0.153
LLAMA B	0.702	0.383	0.046
LLAMA F	0.681	-0.057	-0.113
FDS	0.100	0.880	0.076
OSPAN	0.106	0.743	-0.320
SRT	-0.047	0.036	0.881
LLAMA D	0.211	0.373	-0.558

Table E: Descriptive statistics for the forward digit span and operation span tasks

Test	n	k	M		95% CI	SD		Skew	S-W (p)	α
			Raw	Cor.		Raw	Cor.			
FDS	81	28	131.48	78.40	75.26 – 84.54	23.80	14.20	-0.30	.013	0.86

OSPAN	83	18	84.14	84.99	82.58 – 87.41	10.93	11.04	-1.19	.000	0.73

Note. n – number of participants; k – number of items

Table F: Descriptive statistics for gap-fill, Reading, and Listening

	M		95% CI	SD		Skew	S-W (p)	α
	Raw	Cor.		Raw	Cor.			
GAP (n = 82)								
Articles	15.46	61.85	59.58 – 64.13	2.59	10.35	-0.831	0.0001	0.98
Past	19.88	79.51	76.67 – 82.36	3.24	12.95	-0.571	0.007	0.99
Passive	22.17	88.68	85.77 – 91.60	3.32	13.27	-1.63	0.0001	0.99
Total	57.51	76.68	74.43 – 78.94	7.70	10.26	-0.898	0.0001	0.99
Oxford Placement Test (n = 77)								
Reading	101.52	84.60	81.82 – 87.38	14.69	12.24	-0.939	0.0001	
Listening	85.78	71.43	68.59 – 74.27	15.02	12.51	-0.209	0.037	
Total	93.56	77.97	75.46 – 80.47	13.23	11.02	-0.621	0.016	

Note. Reliability coefficients unavailable for Oxford Placement test

Appendix 3: R packages used in the study

- Correlational analyses were conducted using the ggpairs() function from the *Ggally* package (Schloerke et al., 2024) in R, version 2.2.1.
- Multilevel modeling was performed with the lmer() function from the *lme4* R package (Bates et al., 2015), version 1.1-35.3, while summary tables and p-values for the models were generated using the *lmerTest* package (Kuznetsova et al., 2017), version 3.1-3.
- Interaction analyses were executed with the interact_plot() and sim_slopes() functions from the *interactions* package for R (Long, 2024), version 1.1.5.
- Missing data were treated using the *mice* package for R (van Buuren & Groothuis-Oudshoorn, 2011), version 3.16.0.
- Other functions in the analysis were part of the R core package (R Core Team, 2021), version 2021.09.2, and (R Core Team, 2024), version 4.3.2.

Appendix 4

Table A: Descriptive statistics for the outcome measures

Measure	n	k	M		SD		Skew	S-W	α
			Raw	Cor.	Raw	Cor.		(p)	
SPR	83	72	10.80[1]		27.35[1]		-0.01	.863	.97[2] – .99[3]
EI	82	72	53.40	74.17	12.14	16.86	-1.41	.001	.81
GAP	82	75	57.51	76.68	7.70	10.26	-0.90	.001	.99

Note. n – number of participants; k – number of items; [1] milliseconds; [2] reliability index for list 1; [3] reliability index for list 2

Table B: Factor loadings of complexity, accuracy, and fluency measures

Measure	Component 1
Lexical complexity	.711
Morphosyntactic complexity	.592
Lexical accuracy	.748
Morphosyntactic accuracy	.673
Fluency	.881

Appendix 5

Table A: Multilevel model 1 (Chapter 5)

```
Linear mixed model fit by maximum likelihood. t-tests use
Satterthwaite's method ['lmerModLmerTest']
Formula: SPR ~ L2PO * ELA + L2PO * WM + L2PO * ILA + (1 + ELA |
Participant) + (1 + ILA | Participant) + (1 + WM | Participant)
   Data: df

     AIC      BIC   logLik deviance df.resid
   730.4    793.7   -347.2    694.4      231

Scaled residuals:
    Min      1Q  Median      3Q     Max
-5.0104 -0.6278  0.0377  0.5410  3.2409

Random effects:
 Groups        Name        Variance  Std.Dev.  Corr
 Participant   (Intercept) 1.969e-09 4.437e-05
               ELA         3.349e-10 1.830e-05 1.00
 Participant.1 (Intercept) 6.748e-08 2.598e-04
               ILA         8.573e-08 2.928e-04 -1.00
 Participant.2 (Intercept) 7.546e-10 2.747e-05
               WM          2.146e-10 1.465e-05 -1.00
 Residual                  9.519e-01 9.757e-01
Number of obs: 249, groups:  Participant, 83

Fixed effects:
              Estimate Std. Error         df t value Pr(>|t|)
(Intercept)   0.011119   0.063034 248.976672   0.176    0.860
L2PO          0.161670   0.065983 248.979290   2.450    0.015 *
ELA           0.092840   0.064556 248.984744   1.438    0.152
WM           -0.014008   0.064477 248.996580  -0.217    0.828
ILA           0.030885   0.079349 248.903218   0.389    0.697
L2PO:ELA     -0.045903   0.060871 248.997357  -0.754    0.451
L2PO:WM       0.004753   0.057733 248.958301   0.082    0.934
L2PO:ILA      0.013585   0.064008 248.904693   0.212    0.832
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Correlation of Fixed Effects:
         (Intr) L2PO   ELA    WM     ILA    L2PO:E L2PO:W
L2PO     -0.006
ELA       0.026 -0.174
WM        0.025 -0.029 -0.159
ILA      -0.047  0.040  0.015  0.167
L2PO:ELA -0.181  0.171 -0.151 -0.152 -0.035
L2PO:WM   0.015 -0.298 -0.075  0.076 -0.078 -0.223
L2PO:ILA  0.027 -0.054 -0.021 -0.056  0.026 -0.054  0.153
optimizer (nloptwrap) convergence code: 0 (OK)
boundary (singular) fit: see help('isSingular')
```

Table B: Multilevel model 2 (Chapter 5)

```
Linear mixed model fit by maximum likelihood. t-tests use
Satterthwaite's method ['lmerModLmerTest']
Formula: EI ~ L2PO * ELA + L2PO * ILA + L2PO * WM + (1 + ELA |
Participant) + (1 + ILA | Participant) + (1 + WM | Participant)
   Data: df

     AIC      BIC   logLik deviance df.resid
  2044.3   2107.7  -1004.2   2008.3      231

Scaled residuals:
    Min      1Q  Median      3Q     Max
-2.9137 -0.6196  0.1552  0.6817  1.9570

Random effects:
 Groups        Name        Variance  Std.Dev.  Corr
 Participant   (Intercept) 2.645e+01 5.143e+00
               ELA         4.534e+00 2.129e+00 -1.00
 Participant.1 (Intercept) 1.103e-05 3.322e-03
               ILA         7.295e-07 8.541e-04 -0.82
 Participant.2 (Intercept) 1.056e-05 3.249e-03
               WM          3.759e-06 1.939e-03 -0.34
 Residual                  1.618e+02 1.272e+01
Number of obs: 249, groups:  Participant, 83

Fixed effects:
             Estimate Std. Error        df t value Pr(>|t|)
(Intercept)  74.95420    1.02603  81.74407  73.053  < 2e-16 ***
L2PO          9.99223    1.07396  84.58040   9.304 1.37e-14 ***
ELA           4.81769    1.04798 113.73551   4.597 1.12e-05 ***
ILA           2.11507    1.03648  96.03860   2.041   0.0440 *
WM           -1.81755    1.01460  94.40463  -1.791   0.0764 .
L2PO:ELA     -1.98617    0.99450 117.20821  -1.997   0.0481 *
L2PO:ILA      0.71574    0.90549  91.38259   0.790   0.4313
L2PO:WM       0.04344    0.92630  92.47929   0.047   0.9627
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Correlation of Fixed Effects:
         (Intr) L2PO   ELA    ILA    WM     L2PO:E L2PO:I
L2PO      0.049
ELA      -0.183 -0.176
ILA       0.016 -0.075  0.095
WM        0.029 -0.035 -0.131  0.106
L2PO:ELA -0.163 -0.052 -0.224 -0.109 -0.148
L2PO:ILA -0.031  0.006 -0.012  0.396  0.060 -0.052
L2PO:WM  -0.012 -0.303 -0.042  0.073  0.053 -0.175  0.095
optimizer (nloptwrap) convergence code: 0 (OK)
boundary (singular) fit: see help('isSingular')
```

Table C: Multilevel model 3 (Chapter 5)

```
Linear mixed model fit by maximum likelihood. t-tests use
Satterthwaite's method ['lmerModLmerTest']
Formula: GAP ~ L2PO * ELA + L2PO * ILA + L2PO * WM + (1 + ELA |
Participant) + (1 + ILA | Participant) + (1 + WM | Participant)
   Data: df

     AIC      BIC   logLik deviance df.resid
  1949.9   2013.2   -957.0   1913.9      231

Scaled residuals:
    Min      1Q  Median      3Q     Max
-3.4211 -0.5761  0.0952  0.6768  2.1984

Random effects:
 Groups        Name        Variance  Std.Dev.  Corr
 Participant   (Intercept) 8.341e+00 2.888e+00
               ELA         1.104e+00 1.050e+00 -1.00
 Participant.1 (Intercept) 0.000e+00 0.000e+00
               ILA         5.532e-10 2.352e-05  NaN
 Participant.2 (Intercept) 5.612e+00 2.369e+00
               WM          6.664e-01 8.164e-01 -1.00
 Residual                  1.140e+02 1.068e+01
Number of obs: 249, groups:  Participant, 83

Fixed effects:
            Estimate Std. Error      df t value Pr(>|t|)
(Intercept)  82.0263     0.8181 85.4037 100.265  < 2e-16 ***
L2PO          6.3370     0.8671 47.2888   7.308  2.7e-09 ***
ELA           3.3611     0.8451 87.4654   3.977 0.000143 ***
ILA           0.2768     0.8533 58.4405   0.324 0.746819
WM           -2.1904     0.8557 11.5733  -2.560 0.025625 *
L2PO:ELA     -1.0380     0.7950 70.2267  -1.306 0.195920
L2PO:ILA      0.3427     0.7223 44.8801   0.474 0.637487
L2PO:WM       1.1805     0.7533 53.4256   1.567 0.122983
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Correlation of Fixed Effects:
         (Intr) L2PO   ELA    ILA    WM     L2PO:E L2PO:I
L2PO      0.023
ELA      -0.062 -0.176
ILA       0.022 -0.061  0.091
WM       -0.038 -0.060 -0.153  0.072
L2PO:ELA -0.174  0.073 -0.193 -0.094 -0.166
L2PO:ILA -0.030  0.009 -0.025  0.356  0.015 -0.022
L2PO:WM  -0.010 -0.361 -0.067  0.035  0.123 -0.201  0.087
optimizer (nloptwrap) convergence code: 0 (OK)
boundary (singular) fit: see help('isSingular')
```

Appendix 6: Simple slopes analysis

Output A: Interaction between speaking proficiency and relationship between explicit aptitude and self-paced reading scores

```
> sim_slopes(model.Lmer, pred = ELA, modx = L2PO, jnplot = T,
control.fdr = F, modx.values = c(-2, -1, 0, 1, 2))

JOHNSON-NEYMAN INTERVAL

The Johnson-Neyman interval could not be found. Is the p value for your
interaction term below the specified
alpha?

SIMPLE SLOPES ANALYSIS

Slope of ELA when L2PO = -2.00:

  Est.   S.E.   t val.      p
------ ------ -------- ------
  0.14   0.15     0.91   0.36

Slope of ELA when L2PO = -1.00:

  Est.   S.E.   t val.      p
------ ------ -------- ------
  0.13   0.10     1.30   0.20

Slope of ELA when L2PO =  0.00:

  Est.   S.E.   t val.      p
------ ------ -------- ------
  0.12   0.07     1.78   0.08

Slope of ELA when L2PO =  1.00:

  Est.   S.E.   t val.      p
------ ------ -------- ------
  0.11   0.09     1.28   0.20

Slope of ELA when L2PO =  2.00:

  Est.   S.E.   t val.      p
------ ------ -------- ------
  0.10   0.13     0.73   0.47
```

Output B: Interaction between speaking proficiency and relationship between implicit aptitude and self-paced reading scores

```
> sim_slopes(model.Lmer, pred = ILA, modx = L2PO, jnplot = T,
control.fdr = F, modx.values = c(-2, -1, 0, 1, 2))

JOHNSON-NEYMAN INTERVAL

The Johnson-Neyman interval could not be found. Is the p value for your
interaction term below the specified
alpha?

SIMPLE SLOPES ANALYSIS

Slope of ILA when L2PO = -2.00:

  Est.   S.E.   t val.      p
------ ------ -------- ------
  0.00   0.12     0.03   0.97

Slope of ILA when L2PO = -1.00:

  Est.   S.E.   t val.      p
------ ------ -------- ------
  0.01   0.07     0.14   0.89

Slope of ILA when L2PO =  0.00:

  Est.   S.E.   t val.      p
------ ------ -------- ------
  0.02   0.07     0.25   0.80

Slope of ILA when L2PO =  1.00:

  Est.   S.E.   t val.      p
------ ------ -------- ------
  0.02   0.11     0.22   0.83

Slope of ILA when L2PO =  2.00:

  Est.   S.E.   t val.      p
------ ------ -------- ------
  0.03   0.16     0.19   0.85
```

Output C: Interaction between speaking proficiency and relationship between working memory and self-paced reading scores

```
> sim_slopes(model.Lmer, pred = WM, modx = L2PO, jnplot = T,
control.fdr = F, modx.values = c(-2, -1, 0, 1, 2))

JOHNSON-NEYMAN INTERVAL

The Johnson-Neyman interval could not be found. Is the p value for your
interaction term below the specified
alpha?

SIMPLE SLOPES ANALYSIS

Slope of WM when L2PO = -2.00:

   Est.   S.E.   t val.      p
------- ------ -------- ------
  -0.08   0.13    -0.58   0.56

Slope of WM when L2PO = -1.00:

   Est.   S.E.   t val.      p
------- ------ -------- ------
  -0.06   0.09    -0.74   0.46

Slope of WM when L2PO =  0.00:

   Est.   S.E.   t val.      p
------- ------ -------- ------
  -0.05   0.07    -0.74   0.46

Slope of WM when L2PO =  1.00:

   Est.   S.E.   t val.      p
------- ------ -------- ------
  -0.04   0.10    -0.37   0.71

Slope of WM when L2PO =  2.00:

   Est.   S.E.   t val.      p
------- ------ -------- ------
  -0.02   0.15    -0.15   0.88
```

Output D: Interaction between speaking proficiency and relationship between explicit aptitude and elicited imitation scores

```
> sim_slopes(model.Lmer, pred = ELA, modx = L2PO, jnplot = T,
control.fdr = F, modx.values = c(-2, -1, 0, 1, 2))

When L2PO is INSIDE the interval [-22.15, 1.09], the slope of ELA is p
< .05.

Note: The range of observed values of L2PO is [-2.89, 2.22]

SIMPLE SLOPES ANALYSIS

Slope of ELA when L2PO = -2.00:

  Est.   S.E.   t val.      p
------ ------ -------- ------
  8.20   2.47     3.32   0.00

Slope of ELA when L2PO = -1.00:

  Est.   S.E.   t val.      p
------ ------ -------- ------
  6.41   1.61     3.98   0.00

Slope of ELA when L2PO =  0.00:

  Est.   S.E.   t val.      p
------ ------ -------- ------
  4.61   1.05     4.37   0.00

Slope of ELA when L2PO =  1.00:

  Est.   S.E.   t val.      p
------ ------ -------- ------
  2.81   1.29     2.18   0.03

Slope of ELA when L2PO =  2.00:

  Est.   S.E.   t val.      p
------ ------ -------- ------
  1.01   2.06     0.49   0.62
```

Output E: Interaction between speaking proficiency and relationship between implicit aptitude and elicited imitation scores

```
> sim_slopes(model.Lmer, pred = ILA, modx = L2PO, jnplot = T,
control.fdr = F, modx.values = c(-2, -1, 0, 1, 2))

JOHNSON-NEYMAN INTERVAL

When L2PO is INSIDE the interval [-0.31, 0.36], the slope of ILA is p <
.05.

Note: The range of observed values of L2PO is [-2.89, 2.22]

SIMPLE SLOPES ANALYSIS

Slope of ILA when L2PO = -2.00:

  Est.   S.E.   t val.      p
------ ------ -------- ------
  0.68   1.69     0.40   0.69

Slope of ILA when L2PO = -1.00:

  Est.   S.E.   t val.      p
------ ------ -------- ------
  1.40   1.07     1.30   0.20

Slope of ILA when L2PO =  0.00:

  Est.   S.E.   t val.      p
------ ------ -------- ------
  2.12   1.04     2.04   0.04

Slope of ILA when L2PO =  1.00:

  Est.   S.E.   t val.      p
------ ------ -------- ------
  2.83   1.62     1.74   0.08

Slope of ILA when L2PO =  2.00:

  Est.   S.E.   t val.      p
------ ------ -------- ------
  3.55   2.42     1.47   0.15
```

Output F: Interaction between speaking proficiency and relationship between working memory and elicited imitation scores

```
> sim_slopes(model.Lmer, pred = WM, modx = L2PO, jnplot = T,
control.fdr = F, modx.values = c(-2, -1, 0, 1, 2))

JOHNSON-NEYMAN INTERVAL

The Johnson-Neyman interval could not be found. Is the p value for your
interaction term below the specified
alpha?

SIMPLE SLOPES ANALYSIS

Slope of WM when L2PO = -2.00:

   Est.   S.E.   t val.      p
------- ------ -------- ------
  -1.90   2.06    -0.92   0.36

Slope of WM when L2PO = -1.00:

   Est.   S.E.   t val.      p
------- ------ -------- ------
  -1.86   1.34    -1.39   0.17

Slope of WM when L2PO =  0.00:

   Est.   S.E.   t val.      p
------- ------ -------- ------
  -1.82   1.01    -1.79   0.08

Slope of WM when L2PO =  1.00:

   Est.   S.E.   t val.      p
------- ------ -------- ------
  -1.77   1.41    -1.26   0.21

Slope of WM when L2PO =  2.00:

   Est.   S.E.   t val.      p
------- ------ -------- ------
  -1.73   2.16    -0.80   0.42
```

Output G: Interaction between speaking proficiency and relationship between explicit aptitude and gap-fill scores

```
> sim_slopes(model.Lmer, pred = ELA, modx = L2PO, jnplot = T,
control.fdr = F, modx.values = c(-2, -1, 0, 1, 2))

JOHNSON-NEYMAN INTERVAL

When L2PO is INSIDE the interval [-5.47, 1.13], the slope of ELA is p <
.05.

Note: The range of observed values of L2PO is [-2.89, 2.22]

SIMPLE SLOPES ANALYSIS

Slope of ELA when L2PO = -2.00:

  Est.   S.E.   t val.      p
------ ------ -------- ------
  5.44   1.94     2.80   0.01

Slope of ELA when L2PO = -1.00:

  Est.   S.E.   t val.      p
------ ------ -------- ------
  4.40   1.27     3.47   0.00

Slope of ELA when L2PO =  0.00:

  Est.   S.E.   t val.      p
------ ------ -------- ------
  3.36   0.85     3.98   0.00

Slope of ELA when L2PO =  1.00:

  Est.   S.E.   t val.      p
------ ------ -------- ------
  2.32   1.04     2.23   0.03

Slope of ELA when L2PO =  2.00:

  Est.   S.E.   t val.      p
------ ------ -------- ------
  1.29   1.65     0.78   0.44
```

Output H: Interaction between speaking proficiency and relationship between implicit aptitude and gap-fill scores

```
> sim_slopes(model.Lmer, pred = ILA, modx = L2PO, jnplot = T,
control.fdr = F, modx.values = c(-2, -1, 0, 1, 2))

JOHNSON-NEYMAN INTERVAL

The Johnson-Neyman interval could not be found. Is the p value for your
interaction term below the specified
alpha?

SIMPLE SLOPES ANALYSIS

Slope of ILA when L2PO = -2.00:

   Est.   S.E.   t val.      p
------- ------ -------- ------
  -0.41   1.39    -0.29   0.77

Slope of ILA when L2PO = -1.00:

   Est.   S.E.   t val.      p
------- ------ -------- ------
  -0.07   0.90    -0.07   0.94

Slope of ILA when L2PO =  0.00:

  Est.   S.E.   t val.      p
------ ------ -------- ------
  0.28   0.85     0.32   0.75

Slope of ILA when L2PO =  1.00:

  Est.   S.E.   t val.      p
------ ------ -------- ------
  0.62   1.30     0.48   0.64

Slope of ILA when L2PO =  2.00:

  Est.   S.E.   t val.      p
------ ------ -------- ------
  0.96   1.92     0.50   0.62
```

Output I: Interaction between speaking proficiency and relationship between working memory and gap-fill scores

```
> sim_slopes(model.Lmer, pred = WM, modx = L2PO, jnplot = T,
control.fdr = F, modx.values = c(-2, -1, 0, 1, 2))

JOHNSON-NEYMAN INTERVAL

When L2PO is INSIDE the interval [-7.48, 0.32], the slope of WM is p <
.05.

Note: The range of observed values of L2PO is [-2.89, 2.22]

SIMPLE SLOPES ANALYSIS

Slope of WM when L2PO = -2.00:

   Est.   S.E.   t val.      p
------- ------ -------- ------
  -4.55   1.64    -2.78   0.01

Slope of WM when L2PO = -1.00:

   Est.   S.E.   t val.      p
------- ------ -------- ------
  -3.37   1.07    -3.16   0.00

Slope of WM when L2PO =  0.00:

   Est.   S.E.   t val.      p
------- ------ -------- ------
  -2.19   0.86    -2.56   0.03

Slope of WM when L2PO =  1.00:

   Est.   S.E.   t val.      p
------- ------ -------- ------
  -1.01   1.21    -0.84   0.43

Slope of WM when L2PO =  2.00:

  Est.   S.E.   t val.      p
------ ------ -------- ------
  0.17   1.82     0.09   0.93
```

Appendix 7

Table A: Factor loadings of cognitive variables

	Components	
Measure	1	2
LLAMA B	0.677	-0.104
LLAMA D	0.211	-0.674
LLAMA E	0.780	-0.23
LLAMA F	0.701	0.091
SRT	0.081	0.812

Table B: Factor loadings of L2 knowledge variables

	Components	
Measure	1	2
Self-paced reading	0.283	1.000
Elicited imitation	0.910	0.240
Gap-fill	0.907	0.275

Appendix 8

Table A: Multilevel model 1 (Chapter 6)

```
Linear mixed model fit by maximum likelihood . t-tests use
Satterthwaite's method ['lmerModLmerTest']
Formula: SPR ~ L2PO + L2PR + L2PL + L2PO * ELA + L2PR * ELA + L2PL *
ELA + L2PO * ILA + L2PR * ILA + L2PL * ILA + (1 + ELA | Participant) +
(1 + ILA | Participant)
   Data: df

     AIC      BIC   logLik deviance df.resid
   721.8    788.7   -341.9    683.8      230

Scaled residuals:
    Min      1Q  Median      3Q     Max
-5.2858 -0.6425  0.0468  0.6379  3.1418

Random effects:
 Groups        Name        Variance  Std.Dev.  Corr
 Participant   (Intercept) 2.048e-10 1.431e-05
               ELA         6.570e-12 2.563e-06 -1.00
 Participant.1 (Intercept) 0.000e+00 0.000e+00
               ILA         1.918e-11 4.379e-06  NaN
 Residual                  9.125e-01 9.552e-01
Number of obs: 249, groups:  Participant, 83

Fixed effects:
              Estimate Std. Error         df t value Pr(>|t|)
(Intercept)   0.035452   0.065898 248.999997   0.538 0.591068
L2PO          0.270368   0.076909 248.999997   3.515 0.000521 ***
L2PR         -0.285876   0.102007 248.999997  -2.803 0.005469 **
L2PL          0.060712   0.078690 248.999997   0.772 0.441123
ELA           0.193446   0.078634 248.999996   2.460 0.014571 *
ILA           0.076867   0.074887 248.999988   1.026 0.305679
L2PO:ELA      0.005406   0.069164 248.999993   0.078 0.937760
L2PR:ELA     -0.116409   0.108122 248.999996  -1.077 0.282680
L2PL:ELA      0.025181   0.097290 248.999997   0.259 0.795987
L2PO:ILA      0.055353   0.070583 248.999995   0.784 0.433655
L2PR:ILA     -0.100506   0.104426 248.999995  -0.962 0.336752
L2PL:ILA      0.031866   0.090499 248.999996   0.352 0.725049
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Correlation of Fixed Effects:
         (Intr) L2PO   L2PR   L2PL   ELA    ILA    L2PO:E L2PR:E L2PL:E
L2PO:I L2PR:I
L2PO      0.121
L2PR     -0.195 -0.423
L2PL      0.039 -0.161 -0.442
ELA       0.149  0.123 -0.416  0.002
ILA       0.054  0.138 -0.162 -0.032  0.199
L2PO:ELA  0.022  0.269 -0.317  0.100  0.029  0.068
L2PR:ELA -0.311 -0.388  0.577 -0.066 -0.483 -0.296 -0.409
L2PL:ELA  0.068  0.193 -0.256 -0.071  0.414  0.282 -0.002 -0.609
```

```
L2PO:ILA  0.081  0.266 -0.301  0.048  0.074  0.481  0.233 -0.354  0.135
L2PR:ILA  0.060 -0.137 -0.078  0.172 -0.081 -0.200 -0.153  0.019  0.003
-0.094
L2PL:ILA -0.097 -0.073  0.205 -0.169  0.122 -0.178 -0.088  0.154  0.047
-0.424 -0.554
optimizer (nloptwrap) convergence code: 0 (OK)
boundary (singular) fit: see help('isSingular')
```

Table B: Multilevel model 2 (Chapter 6)

```
Linear mixed model fit by maximum likelihood . t-tests use
Satterthwaite's method ['lmerModLmerTest']
Formula: EI ~ L2PO + L2PR + L2PL + L2PO * ELA + L2PR * ELA + L2PL * ELA
+ L2PO * ILA + L2PR * ILA + L2PL * ILA + (1 + ELA | Participant) + (1 +
ILA | Participant)
   Data: df

     AIC      BIC   logLik deviance df.resid
  2032.3   2099.1   -997.1   1994.3      230

Scaled residuals:
    Min      1Q  Median      3Q     Max
-3.0916 -0.5766  0.0922  0.7043  2.3173

Random effects:
 Groups        Name        Variance Std.Dev. Corr
 Participant   (Intercept)   0.000   0.000
               ELA           1.570   1.253    NaN
 Participant.1 (Intercept)   2.127   1.459
               ILA          23.164   4.813   -1.00
 Residual                  160.271  12.660
Number of obs: 249, groups:  Participant, 83

Fixed effects:
            Estimate Std. Error       df t value Pr(>|t|)
(Intercept) 74.81890    1.00286 99.66188  74.605  < 2e-16 ***
L2PO         7.65155    1.20611 80.69074   6.344 1.22e-08 ***
L2PR         5.13357    1.56910 84.92637   3.272  0.00155 **
L2PL         0.38569    1.19930 94.48669   0.322  0.74847
ELA          2.70731    1.19344 53.50016   2.268  0.02736 *
ILA          3.24203    1.38650 39.54393   2.338  0.02452 *
L2PO:ELA    -1.40056    1.05383 17.06215  -1.329  0.20134
L2PR:ELA    -0.47332    1.61775 50.53531  -0.293  0.77104
L2PL:ELA    -0.10776    1.56715 54.76118  -0.069  0.94543
L2PO:ILA     1.77042    1.31620 21.84450   1.345  0.19239
L2PR:ILA    -0.05444    1.77595 41.84767  -0.031  0.97569
L2PL:ILA    -2.35082    1.61704 27.41973  -1.454  0.15736
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Correlation of Fixed Effects:
         (Intr) L2PO   L2PR   L2PL   ELA    ILA    L2PO:E L2PR:E L2PL:E
L2PO:I L2PR:I
L2PO      0.088
L2PR     -0.164 -0.445
L2PL      0.044 -0.129 -0.450
ELA       0.141  0.123 -0.432  0.034
ILA      -0.112  0.111 -0.071 -0.074  0.157
L2PO:ELA  0.027  0.253 -0.293  0.096 -0.012  0.023
L2PR:ELA -0.314 -0.394  0.572 -0.082 -0.449 -0.200 -0.418
L2PL:ELA  0.045  0.213 -0.271 -0.021  0.414  0.257 -0.005 -0.597
L2PO:ILA  0.111  0.086 -0.194  0.052  0.019  0.216  0.233 -0.301  0.052
L2PR:ILA  0.069 -0.027 -0.237  0.235  0.008 -0.209 -0.139 -0.025  0.046
-0.122
```

```
L2PL:ILA -0.111 -0.090  0.274 -0.214  0.059 -0.035 -0.120  0.195 -0.003
-0.390 -0.522
optimizer (nloptwrap) convergence code: 0 (OK)
boundary (singular) fit: see help('isSingular')
```

Table C: Multilevel model 3 (Chapter 6)

```
Linear mixed model fit by maximum likelihood . t-tests use
Satterthwaite's method ['lmerModLmerTest']
Formula: GAP ~ L2PO + L2PR + L2PL + L2PO * ELA + L2PR * ELA + L2PL *
ELA + L2PO * ILA + L2PR * ILA + L2PL * ILA + (1 + ELA | Participant) +
(1 + ILA | Participant)
   Data: df

     AIC      BIC   logLik deviance df.resid
  1950.8   2017.7   -956.4   1912.8      230

Scaled residuals:
    Min      1Q  Median      3Q     Max
-3.4881 -0.5279  0.1003  0.6550  2.1867

Random effects:
 Groups        Name        Variance  Std.Dev.  Corr
 Participant   (Intercept) 1.521e+01 3.900e+00
               ELA         4.311e+00 2.076e+00 -1.00
 Participant.1 (Intercept) 0.000e+00 0.000e+00
               ILA         1.395e-10 1.181e-05  NaN
 Residual                  1.118e+02 1.057e+01
Number of obs: 249, groups:  Participant, 83

Fixed effects:
            Estimate Std. Error       df t value Pr(>|t|)
(Intercept)  81.6262     0.8865  87.6392  92.072  < 2e-16 ***
L2PO          4.8722     1.0263  92.7761   4.747 7.47e-06 ***
L2PR          1.5038     1.3354 112.5656   1.126   0.2625
L2PL          2.2640     1.0791  76.2385   2.098   0.0392 *
ELA           1.8072     1.0804  95.9672   1.673   0.0976 .
ILA          -0.3806     0.9877 110.0668  -0.385   0.7007
L2PO:ELA     -1.8423     1.0133  56.4376  -1.818   0.0744 .
L2PR:ELA      2.3416     1.4926  95.1924   1.569   0.1200
L2PL:ELA     -1.0162     1.2832 131.9355  -0.792   0.4298
L2PO:ILA     -0.5472     0.9307 105.5583  -0.588   0.5578
L2PR:ILA      2.7723     1.4325  84.0581   1.935   0.0563 .
L2PL:ILA     -0.9884     1.1999 122.7799  -0.824   0.4117
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Correlation of Fixed Effects:
         (Intr) L2PO   L2PR   L2PL   ELA    ILA    L2PO:E L2PR:E L2PL:E
L2PO:I L2PR:I
L2PO      0.134
L2PR     -0.131 -0.372
L2PL      0.014 -0.201 -0.451
ELA      -0.048  0.120 -0.407  0.001
ILA       0.063  0.161 -0.195 -0.031  0.235
L2PO:ELA  0.062  0.026 -0.273  0.175 -0.014  0.084
L2PR:ELA -0.303 -0.316  0.435  0.001 -0.439 -0.334 -0.463
L2PL:ELA  0.078  0.217 -0.175 -0.250  0.394  0.310 -0.011 -0.595
L2PO:ILA  0.071  0.267 -0.305  0.042  0.110  0.518  0.180 -0.326  0.126
L2PR:ILA  0.048 -0.160 -0.082  0.177 -0.103 -0.214 -0.035  0.000  0.018
-0.131
```

```
L2PL:ILA -0.103 -0.064  0.220 -0.184  0.145 -0.181 -0.179  0.180  0.068
-0.418 -0.540
optimizer (nloptwrap) convergence code: 0 (OK)
boundary (singular) fit: see help('isSingular')
```

Appendix 9

Table A: Multilevel model 1 (Chapter 7)

```
Linear mixed model fit by maximum likelihood . t-tests use
Satterthwaite's method ['lmerModLmerTest']
Formula: SPR ~ Structure * ELA + Structure * ILA + Structure * WM + ELA
+ ILA + WM + (1 + ELA | Participant) + (1 + ILA | Participant) + (1 +
WM | Participant)
   Data: df

     AIC      BIC   logLik deviance df.resid
   729.8    807.1   -342.9    685.8      227

Scaled residuals:
    Min      1Q  Median      3Q     Max
-4.1802 -0.6304 -0.0616  0.4740  3.9367

Random effects:
 Groups        Name        Variance  Std.Dev.  Corr
 Participant   (Intercept) 0.000e+00 0.000e+00
               ELA         8.835e-10 2.972e-05  NaN
 Participant.1 (Intercept) 4.868e-08 2.206e-04
               ILA         8.567e-09 9.256e-05 1.00
 Participant.2 (Intercept) 5.287e-02 2.299e-01
               WM          2.020e-02 1.421e-01 1.00
 Residual                  8.566e-01 9.255e-01
Number of obs: 249, groups:  Participant, 83

Fixed effects:
                 Estimate Std. Error         df t value Pr(>|t|)
(Intercept)     -0.123491   0.105737 245.131868  -1.168   0.2440
Structure2       0.089467   0.143710 193.444117   0.623   0.5343
Structure3       0.278046   0.143710 193.444117   1.935   0.0545 .
ELA              0.054253   0.106153 247.804230   0.511   0.6097
ILA              0.056938   0.104564 240.915944   0.545   0.5866
WM              -0.243936   0.108121 175.527438  -2.256   0.0253 *
Structure2:ELA   0.143359   0.144934 193.444117   0.989   0.3238
Structure3:ELA   0.121873   0.144934 193.444117   0.841   0.4014
Structure2:ILA  -0.006169   0.143987 193.444117  -0.043   0.9659
Structure3:ILA  -0.101467   0.143987 193.444117  -0.705   0.4818
Structure2:WM    0.294635   0.145734 193.444117   2.022   0.0446 *
Structure3:WM    0.294876   0.145734 193.444117   2.023   0.0444 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Correlation of Fixed Effects:
            (Intr) Strct2 Strct3 ELA    ILA    WM     S2:ELA S3:ELA
S2:ILA S3:ILA St2:WM
Structure2  -0.680
Structure3  -0.680  0.500
ELA          0.003 -0.001 -0.001
ILA          0.023 -0.014 -0.014  0.061
WM           0.063 -0.008 -0.008 -0.187  0.070
Strctr2:ELA -0.001  0.001  0.000 -0.683 -0.046  0.124
```

```
Strctr3:ELA -0.001  0.000  0.001 -0.683 -0.046  0.124  0.500
Strctr2:ILA -0.014  0.020  0.010 -0.045 -0.689 -0.046  0.066  0.033
Strctr3:ILA -0.014  0.010  0.020 -0.045 -0.689 -0.046  0.033  0.066
0.500
Structr2:WM -0.008  0.011  0.006  0.126 -0.047 -0.674 -0.185 -0.092
0.069  0.034
Structr3:WM -0.008  0.006  0.011  0.126 -0.047 -0.674 -0.092 -0.185
0.034  0.069  0.500
optimizer (nloptwrap) convergence code: 0 (OK)
boundary (singular) fit: see help('isSingular')
```

Table B: Multilevel model 2 (Chapter 7)

```
Linear mixed model fit by maximum likelihood . t-tests use
Satterthwaite's method ['lmerModLmerTest']
Formula: EI ~ Structure * ELA + Structure * ILA + Structure * WM + ELA
+ ILA + WM + (1 + ELA | Participant) + (1 + ILA | Participant) + (1 +
WM | Participant)
   Data: df

     AIC      BIC   logLik deviance df.resid
  1963.1   2040.5   -959.5   1919.1      227

Scaled residuals:
    Min      1Q  Median      3Q     Max
-3.7089 -0.5444  0.0412  0.5868  1.8954

Random effects:
 Groups        Name        Variance Std.Dev. Corr
 Participant   (Intercept) 119.239  10.920
               ELA          16.825   4.102   -1.00
 Participant.1 (Intercept)   9.354   3.058
               ILA           2.562   1.601   1.00
 Participant.2 (Intercept)  13.098   3.619
               WM            6.855   2.618   1.00
 Residual                   66.057   8.128
Number of obs: 249, groups:  Participant, 83

Fixed effects:
               Estimate Std. Error       df t value Pr(>|t|)
(Intercept)     80.4624     1.6704  97.2418  48.169  < 2e-16 ***
Structure2      -1.1969     1.2620 167.8194  -0.948 0.344277
Structure3     -17.4570     1.2620 167.8194 -13.833  < 2e-16 ***
ELA              5.2237     1.4935  86.0199   3.498 0.000746 ***
ILA              2.4259     1.4774   4.7673   1.642 0.164355
WM              -1.5378     1.6043  51.8051  -0.959 0.342218
Structure2:ELA   1.4133     1.2727 167.8194   1.110 0.268405
Structure3:ELA   2.5864     1.2727 167.8194   2.032 0.043708 *
Structure2:ILA  -0.3376     1.2644 167.8194  -0.267 0.789821
Structure3:ILA   0.2835     1.2644 167.8194   0.224 0.822851
Structure2:WM    1.2242     1.2797 167.8194   0.957 0.340161
Structure3:WM    0.3499     1.2797 167.8194   0.273 0.784888
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Correlation of Fixed Effects:
            (Intr) Strct2 Strct3 ELA    ILA    WM     S2:ELA S3:ELA
S2:ILA S3:ILA St2:WM
Structure2  -0.378
Structure3  -0.378  0.500
ELA         -0.364  0.000  0.000
ILA          0.045 -0.009 -0.009  0.057
WM           0.066 -0.005 -0.005 -0.145  0.088
Strctr2:ELA  0.000  0.001  0.000 -0.426 -0.028  0.074
Strctr3:ELA  0.000  0.000  0.001 -0.426 -0.028  0.074  0.500
Strctr2:ILA -0.008  0.020  0.010 -0.028 -0.428 -0.027  0.066  0.033
Strctr3:ILA -0.008  0.010  0.020 -0.028 -0.428 -0.027  0.033  0.066
0.500
```

```
Structr2:WM -0.004  0.011  0.006  0.079 -0.029 -0.399 -0.185 -0.092
0.069  0.034
Structr3:WM -0.004  0.006  0.011  0.079 -0.029 -0.399 -0.092 -0.185
0.034  0.069  0.500
optimizer (nloptwrap) convergence code: 0 (OK)
boundary (singular) fit: see help('isSingular')
```

Table C: Multilevel model 3 (Chapter 7)

```
Linear mixed model fit by maximum likelihood . t-tests use
Satterthwaite's method ['lmerModLmerTest']
Formula: GAP ~ Structure * ELA + Structure * ILA + Structure * WM + ELA
+ ILA + WM + (1 + ELA | Participant) + (1 + ILA | Participant) + (1 +
WM | Participant)
   Data: df

     AIC      BIC   logLik deviance df.resid
  1936.7   2014.1   -946.4   1892.7      227

Scaled residuals:
    Min      1Q  Median      3Q     Max
-3.5418 -0.4849  0.1100  0.5286  2.4316

Random effects:
 Groups        Name        Variance  Std.Dev.  Corr
 Participant   (Intercept) 5.728e+01 7.569e+00
               ELA         8.842e+00 2.974e+00 -1.00
 Participant.1 (Intercept) 0.000e+00 0.000e+00
               ILA         3.514e-09 5.928e-05  NaN
 Participant.2 (Intercept) 7.543e+00 2.746e+00
               WM          4.735e+00 2.176e+00 1.00
 Residual                  7.546e+01 8.687e+00
Number of obs: 249, groups:  Participant, 83

Fixed effects:
              Estimate Std. Error       df t value Pr(>|t|)
(Intercept)     79.3622     1.3562 138.3782  58.518  < 2e-16 ***
Structure2       9.1809     1.3488 169.9396   6.807 1.64e-10 ***
Structure3      -2.5820     1.3488 169.9396  -1.914 0.057266 .
ELA              4.2759     1.2620 186.1915   3.388 0.000859 ***
ILA              1.3133     1.2030 121.9377   1.092 0.277133
WM              -1.2806     1.3351  82.2029  -0.959 0.340265
Structure2:ELA   1.7928     1.3603 169.9396   1.318 0.189292
Structure3:ELA   0.4923     1.3603 169.9396   0.362 0.717901
Structure2:ILA  -0.8291     1.3514 169.9396  -0.613 0.540370
Structure3:ILA  -1.2509     1.3514 169.9396  -0.926 0.355952
Structure2:WM   -1.0347     1.3678 169.9396  -0.756 0.450437
Structure3:WM   -0.4190     1.3678 169.9396  -0.306 0.759746
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Correlation of Fixed Effects:
             (Intr) Strct2 Strct3 ELA    ILA    WM     S2:ELA S3:ELA
S2:ILA S3:ILA St2:WM
Structure2  -0.497
Structure3  -0.497  0.500
ELA         -0.268  0.000  0.000
ILA          0.025 -0.011 -0.011  0.053
WM           0.067 -0.006 -0.006 -0.151  0.071
Strctr2:ELA  0.000  0.001  0.000 -0.539 -0.037  0.095
Strctr3:ELA  0.000  0.000  0.001 -0.539 -0.037  0.095  0.500
Strctr2:ILA -0.010  0.020  0.010 -0.036 -0.562 -0.035  0.066  0.033
Strctr3:ILA -0.010  0.010  0.020 -0.036 -0.562 -0.035  0.033  0.066
0.500
```

```
Structr2:WM -0.006  0.011  0.006  0.099 -0.039 -0.512 -0.185 -0.092
0.069  0.034
Structr3:WM -0.006  0.006  0.011  0.099 -0.039 -0.512 -0.092 -0.185
0.034  0.069  0.500
optimizer (nloptwrap) convergence code: 0 (OK)
boundary (singular) fit: see help('isSingular')
```