« Quelle structure grammaticale est la plus difficile ? » : Foreign Language Aptitude & Working Memory in Second Language Acquisition of French

By Julia Gamache

An Undergraduate Honors Thesis

Linguistics Program

University of Vermont

Defense Committee:

Thesis Advisor: Guillermo Rodríguez, Ph.D.

Committee Chair: Ching Selao, Ph.D.

Committee Member: Ben Sienicki, Ph.D.

Table of Contents

Abstract……3

1. Introduction……4
2. Literature Review……...…4
    2.1. Learner Variables in SLA……...5
    2.2. Aptitude Testing……8
    2.3. Replicating Yalçin & Spada (2016) ……..…10
3. Methods & Procedures……15
    3.1. Target Structures……...15
    3.2. Grammaticality Judgment Tasks……..17
    3.3. Testing of Aptitude & Cognition……..18
4. Results……..19
5. Discussion……22
6. Conclusions……..26

Acknowledgments……..27

References……..28

Appendix A: French GJT Sample Items……30

Appendix B: LLAMA……31

Appendix C: Symmetry Span……35

Appendix D: French Placement Test Scores……36

ABSTRACT

This study investigates the correlations between foreign language aptitude (FLA), working memory capacity (WM) and learning of a second language (L2) in a university classroom setting of native English-speaking learners of elementary level French at the University of Vermont. Proficiency was assessed separately for two unique target structures hypothesized to vary in grammatical difficulty based on complexity of construction and crosslinguistic influence: the near future tense (the easy structure) and partitive articles (the difficult structure). Proficiency was measured using a timed, written grammaticality judgment task (GJT). A pretest was administered prior to exposure to either structure, and posttests were administered in the week following introduction of the relevant material. 23 out of a total of 71 students volunteered for cognitive testing, examining FLA with the LLAMA aptitude test (Meara & Rogers, 2019) and WM using a Symmetry Span task (Draheim et al., 2018). The results did not show significant correlations between FLA and proficiency in either target structure, but positive correlations between FLA and WM as well as between WM and GJT scores for the near future highlight important interactions between cognitive function and successful SLA.

1. INTRODUCTION

Primarily, second language acquisition (SLA) is concerned with one's ability to acquire a language in adulthood, or, at the very least, past the developmental period in which most children acquire their native language through naturalistic means (DeKeyser, 2000). The idea of an innate, largely stable quality determinant of one's ability to acquire a foreign language efficiently and effectively has been of relevance to scholars in the field dating back to the mid twentieth century (Carroll, 1958). Today, second language acquisition (SLA) is frequently thought of as distinct from (but associated with) general cognitive aptitude, such that foreign language aptitude (FLA) is specific to the context of SLA and serves as a predictor of proficiency. Success is not synonymous with aptitude, but rather is the result of learning that takes place which can be aided or facilitated by this "innate" aptitude (Skehan, 2015). Early frameworks, however, centered on FLA as a reflection of the speed and accuracy with which one acquires a second language (Carroll, 1962), that is, if someone acquired proficiency in a second language quickly and with a high degree of accuracy (particularly in morpho-syntax), the assumption was that they must have high FLA. Thus, FLA was thought as a unitary capacity that helped in the acquisition of additional languages. Instead, more recently, scholars have begun to decompose FLA (Ameringer et. al 2018) into potential components that target purely linguistic abilities and/or more general, language-independent, cognitive capacities.

2. LITERATURE REVIEW

In general terms, FLA's components remain quite similar despite the nuances investigated by research. Those abilities, modeled after Carroll's work (1981), encompass phonetic encoding, sensitivity to grammatical structure, rote memory, and inductive language learning capacity (Ameringer et. al, 2018). Many variables have been theorized to affect second

language acquisition in adults, particularly related to the context of language instruction. Research efforts have expanded to examine learner variables like working memory (Mujtaba et al., 2021), motivation (Hyltenstam, 2021), multitasking (Elsmore, 1994), typological distance between the L1 and L2 (Bokander, 2020), and the current focus seems to be in assessing the predictive validity of such factors upon FLA test battery results (Foryś-Nogala, 2021; Mujtaba et al., 2021; Li & Zhao, 2021) and/or performance in the L2 (Rogers et al., 2023; Bokander & Bylund, 2020). Thus, we enter the conversation of how best to assess the nuances of FLA and related learner variables in outcomes that analyze performance or acquisition of novel language in some way (Saito et al., 2019). If there are multiple facets of FLA which may be influenced or 'enacted' variably on biology, affect, etc., robust conclusions can be challenging. Methodology to examine some of these influences has ranged from self-reports of emotional and mental status, perceived ability, IQ matrices tests, motivational guidance from instructors, and more (Bell & McCallum, 2012; Hyltenstam, 2021; Li & Zhao, 2021; Wen, 2021, Sparks et al., 1998). Given this recent interest in the components and assessment of FLA, the following discussion will address individual differences and highlight working memory (WM) as the cognitive capacity that has been claimed to be fundamental in the composition of aptitude.

2.1 Learner Variables in SLA

WM capacity represents the ability to store and process information while completing tasks, and thus is frequently involved in research relating to (especially explicit) L2 learning (Li & Zhao, 2021). Because the existing literature has found that WM can be highly correlated with proficiency and FLA, it has even been hypothesized to act as a suitable equivalent measure to aptitude when it comes to language acquisition processes (Wen, 2016; Wen & Li, 2019). In a study examining the connection between WM and FLA as a predictors of vocabulary acquisition

and writing accuracy, Mujtaba et al. (2021) found that WM did in fact positively correlate with performance in a picture-description writing task. The experiment utilized WM test called the Operation Span Test (OST), structured similarly to the instrument used in this study but which is not language independent as it requires memorization of words as part of the instruction. On the other hand, it has simultaneously been concluded that WM was *not* a significant predictor for proficiency in a study on the role of grammatical inferencing skills and working memory which employed a grammaticality judgment task as well as a reading task (Foryś-Nogala, 2021). Using a digit span task which involves repetition of digits forward and backward, neither grammaticality judgment task nor reading span scores were correlated. Neither of these studies attempted to make distinctions between particular structures when evaluating data. Further, a meta-review conducted by Li & Zhao highlights findings that support the idea of working memory as strongly correlated to but distinct from aptitude (2021). The claim is made on the grounds that aptitude is domain-specific only to language learning, while cognitive capacities such as WM are domain-general. While WM has appeared in literature both as a contributing component to language aptitude and as a standalone cognitive capacity, it is nonetheless necessary to acknowledge it as an epiphenomenon. Involving both storage and processing elements simultaneously, it is responsible for balancing more than one cognitive function at a time, with the subsequent potential to support various language tasks in differing ways (Baddeley, 2017).

Another cognitive variable to explore is attention as a possible predictor of FLA and achievement in a foreign language (Engle, 2018). Existing theory suggests that novel language input, as acquired by adults, typically results from some kind of instruction or guided content that elicits an intentional focus on the material (i.e., focus-on-form/noticing; Long, 2007).

McLaughlin's (1987) cognitive theory, for one, took the stance that learners of second languages have limited capacity for attention to materials at a given time, affecting processing of linguistic input more broadly.

Additionally, some scholars have chosen to focus on how strength of native language proficiency and FLA intertwine, as success in L2 learning may be connected to the metalinguistic knowledge acquired in the first language, together with attitude, anxiety, and motivation (Sparks et al., 1998; 2006). Though attention, motivation, and L1 metalinguistic knowledge are not factors which will be included in the main analyses of this project, it is worthy to keep in mind that these other variables present individual differences and may be determinant of success in the language classroom. It is the diversity of variable to consider that makes formal language instruction environments (whether intact classrooms or lab settings) relatively controlled environment (as compared to naturalistic learners) that researchers use as testing grounds for their hypotheses.

The present study seeks to better understand the interface between grammatical difficulty, WM, and FLA in a foreign language classroom in a university setting. The language classroom has become an excellent resource for scholars to employ as an ecologically valid site for the study of FLA, given that these academic environments often provide a standard application of instructional methods and techniques, and can present a somewhat uniform sample of participants. For example, Saito et al. (2019) found that in a class of Japanese students, high FLA correlated positively with improvement in fluency, pronunciation, and verbal features while learning English over a year. Yalçin & Spada (2016) highlight the purpose of using classroom settings for research purposes: while laboratory environments may retain more control,

classrooms provide a realistic testbed for SLA processes to emerge, while offering a convenient participant pool and standard curricula for these kinds of FLA studies.

2.2 Aptitude Testing

A crucial step in this kind of research is to pick a suitable test to assess FLA. Since the emergence of research on the topic began over sixty years ago, John Carroll's (1959) Modern Language Aptitude Test (MLAT) was one of the first comprehensive batteries, and in fact remains the most widely known and used today. However, the lack of theoretical scaffolding behind his initial attempts at the definition of aptitude have always raised concern in the field, despite significant correlations with L2 learning (Ameringer et al., 2018). The tests which comprise the original MLAT were chosen by administering as many as forty different cognitive tests, identifying which ones matched with measures of language proficiency, and selecting those for the test battery (Carroll, 1959; Ameringer, 2018). Additionally, none of the subtests included in the MLAT accounted for inductive language learning capacity, one of his four theorized components of aptitude. Along with concerns that the test was becoming technologically outdated, scholars were prompted to begin searching for an alternative instrument to assess FLA. The LLAMA battery (Meara, 2005) is formulated using the MLAT as a baseline but following a "language neutrality" principle (Wen, 2021), and is freely available for research use. By contrast, LLAMA's (version 3) online format, lack of use restrictions, and ease of use make it applicable to a broad range of participants, all while having great construct validity as variables like age, gender, education level, and linguistic background appear not to affect the results of this test (Rogers et al., 2017; Ameringer et al., 2018; Wen, 2021). The LLAMA is not suitable for use with children (Rogers et al., 2017) but overall serves as an effective and reliable tool to assess

FLA among adults, and thus it was chosen as the primary data collection instrument for the present study.

The original online battery includes four subtests: LLAMA B (vocabulary learning), LLAMA D (sound recognition), LLAMA E (sound-symbol association), and LLAMA F (grammatical inferencing). LLAMA B includes an exposure phase in which participants have two minutes to memorize the names of twenty foreign objects. The testing phase then prompts the subject with one name at a time, and users must click on the corresponding object. LLAMA D involved a learning phase in which an audio file containing a list of words in a synthetic language was played during the learning phase. Once finished, participants are then played one word at a time, and must select whether the word was novel or had already been played. LLAMA E's exposure phase allows two minutes to interact with 24 buttons that play one syllable, using familiar symbols to represent sounds in unfamiliar ways. The testing phase combines these sounds into twenty buttons each containing the symbols for a two syllable "word". It then plays the audio of each word one a time, participants must click on the matching symbol. Finally, LLAMA F includes 20 buttons which each show an image accompanied by its matching description. After five minutes of learning, testing showed images one at a time and asked subjects to click which of two options was the correct match in a forced-choice format.

Unfortunately, little research yet exists using the recently updated and improved LLAMA v.3. With its update in 2023, Rogers et. al (2023) sought to improve the validity and effectiveness of this test battery in its third version. For example, the procedure for LLAMA F was changed significantly, such that the original test phase which provided participants with a forced-choice selection now prompts actual production of the synthetic language demonstrated in the exposure phase. LLAMA D was also adjusted so that the separate exposure and test phases

were combined into one activity. Instead of listening to the entire collection of exposure words and then having to evaluate data one by one as novel or familiar, participants click a button to play each word and immediately choose whether they have heard it previously in the dataset or not. The first 10 items are all novel, constituting a sort of "learning phase" on their own, but formatted the same as the rest of the activity which follows. LLAMA E's testing phase was also changed from a forced choice format, so that participants must intentionally click one of twenty buttons to match the sound that is played, or simply state that they are unsure, as opposed to guessing with a 50/50 chance of accuracy. Finally, LLAMA B remained largely unchanged aside from cosmetic updates, which the rest of the subtests received as well. The current study provided an opportunity to employ this new version of the LLAMA battery and observe if significant differences may be highlighted in results that could be due to such adjustments.

2.3 Replicating Yalçin & Spada (2016)

Yalçin & Spada (2016), demonstrated in a quasi-experimental, classroom-based study the relationship between foreign language aptitude (FLA) and grammatical difficulty in the acquisition of new structures. A participant pool of 66 adolescent learners of English from Turkey took part in this study. All students received equivalent instruction for four hours on two English structures of varying difficulty: the past progressive as an "easy" structure, and the passive voice as a "difficult" construction. Instruction was not altered for this experiment, though one of the researchers took over as the instructor during these lessons. The level of relative grammatical difficulty was understood as a combination of structural complexity, frequency in the input, time of acquisition for L1 learners, and "perceived difficulty" as assessed through participant ratings. The past progressive "easy" structure was noted as such because of its frequency and accessibility in the input (salience), and clarity of meaning as reflected in its form.

The passive voice was categorized conversely due to the significant number and variety of allomorphs when forming the passive voice, as well as lower frequency and greater complexity of grammatical processes (Yalçin & Spada, 2016, pp. 246-247). This concept of grammatical difficulty becomes relevant when stages of second language acquisition are considered, as research has shown all learners acquire the features of a language in the same general order (Ortega, 2009). What can vary, however, is the time it takes for an individual to progress through each acquisitional phase. Though of course motivation and attention devoted to acquisition vary as well, when receiving equal instruction, all learners are in theory capable of achieving proficiency of a given structure with adequate practice and resources. For a simple structure, there is a greater likelihood that a larger number of learners will successfully and quickly acquire it. Higher FLA has been theorized to give learners an advantage over their peers, and thus discrepancies among the progress of individuals may be highlighted using comparisons between two structures of varying difficulty.

In Yalçin & Spada (2016), knowledge of these two structures was assessed with a written grammaticality judgment task pre- and posttests, in which each section contained 40-items (similar to that in Appendix A). The past progressive GJT contained a variety of uses of the structure, with 21 grammatical and 19 ungrammatical sentences. The passive GJT was created using both the present, past, and present perfect tenses in English, and contained 16 grammatical items and 24 ungrammatical ones. The students had to identify whether each item was correct, incorrect, or mark it as "*I don't know*". If deemed incorrect, they were asked to provide corrections. They were then scored anywhere from zero to three points per item based on their responses and correction attempts made. Productive capacity was additionally measured with two one-on-one oral production tasks (OPT) guided by image cues on a Microsoft PowerPoint

presentation. Each participant also completed aptitude testing with all four sub-tests of the computerized LLAMA battery (Meara, 2005; version 1): vocabulary learning (LLAMA B), sound recognition (LLAMA D), sound-symbol correspondence (LLAMA E) and grammatical inferencing (LLAMA F).

Yalçin & Spada (2016) found that the passive (difficult) structure showed strong correlations between the pretest and posttest scores, as well as between the LLAMA F (grammar induction/inferencing) scores and the GJT posttest scores. Though a significant correlation was also found between the posttest score and LLAMA E (sound-symbol correspondence), LLAMA F was prioritized in the analysis of performance seeing as grammatical inferencing is understood to be the strongest predicting component of FLA (Li, 2016; Foryś-Nogala, 2021). Regression analyses revealed that grammatical inferencing scores predicted 12% of the variability in the Passive GJT posttest. Regarding the passive OPT scores, the only significant correlation was between the pretest and posttest scores; none were observed between scores and predictor variables with the LLAMA battery. Data for the past progressive (easy) structure demonstrated once again a strong correlation between past progressive GJT pretests and posttests. Additionally, LLAMA E (sound-symbol recognition) was associated significantly to the posttest scores for this structure, though the influence of prior knowledge of the structure held by participants resulted in a ceiling effect. After excluding 14 items to adjust for this phenomenon, further analyses showed that aptitude scores correlated with pretest performance, but not with the posttests, differing from the results for the passive structure. OPT results for the past progressive also showed strong correlations between pretests and posttests, such that pretest scores predicted 55% of variation, and, in addition, LLAMA B (vocabulary learning) scores explained a significant amount of this variance.

These results indicate that the participants relied more heavily on aptitude as a resource when facing a more difficult structure, as students with greater aptitude may have picked up the patterns more easily and formulated more hypotheses when engaging with material from explicit instruction. The correlation of LLAMA B (rote/associative memory) scores and OPT scores of the familiar structure may also suggest that memory plays a more significant role in later stages of L2 acquisition, as theorized in the aptitude profile model (Skehan, 2002). This model supports that there are four primary levels of linguistic development: noticing (initial recognition of form), patterning (recognition and manipulation of structures), controlling (proceduralization of use), and lexicalizing (development beyond rule-based processing for real-time usage). Skehan argues that WM becomes more relevant to acquisition in the latter stages, such as when learners begin to recognize and follow patterns in the L2 (patterning), a crucial stage for elementary level learners.

Considering Yalçin & Spada's results on the influence of cognition, in the guise of WM on proficiency, research has led to findings that highlight potential positive correlations between aptitude and WM when it comes to second language performance. New studies tackle the question of how language aptitude differs in connection to implicit vs. explicit learning, as well as among differing experiences in explicit classroom learning, such as type/complexity of the structure being learned. Intact classroom environments thus can prove to be ideal participant pools, as instruction (and prior knowledge, to a degree) is controlled for, minimizing the influence of confounding variables. Thus, cognitive capacities and FLA can be assessed among groups that all receive the same instructional treatment. While their methodology yielded significant results, there remain several implications highlighted for future research. First, their experiment was partially impeded by a ceiling effect in the results for the easy structure, as

explained by the fact that the past progressive was not a novel structure to the participant pool. Yalçin & Spada, like much of the existing studies on FLA, also target English as the L2 in their procedures. Conversely, French has rarely been used in this type of empirical study and provided a unique opportunity for exploration. The latest version of LLAMA (version 3) was released following the publication of their work, and so the development of an improved test battery alongside the addition of a WM test which operates completely independently from language skills allows for partial replication of methodology with the goal to uncover further correlations between FLA, WM, and proficiency when faced with two structures of varying difficulty. Thus, Y&S's procedures were adjusted to accommodate a different population, target language and newer versions of FLA and cognitive tests in this replication. Consequently, this thesis project addresses three central research questions:

R1) *Can FLA scores as measured by LLAMA v.3 successfully predict L2 proficiency of students in an elementary level French course?*

R2) *Do students rely more heavily on their FLA capacity when tasked with learning a more challenging structure, as opposed to a simpler one?*

R3) *What is the relationship of working memory with FLA and with proficiency?*

We hypothesized that higher aptitude scores will correlate with higher proficiency scores of the participants in this study. It is also expected that students with higher aptitude would perform better on the GJT for the difficult structure (partitive articles) than their peers with lower aptitude, even if those lower-aptitude participants performed well on the GJT for the easy structure (the near future). Working memory scores were expected to correlate with both aptitude and proficiency measures.

3. METHODS & PROCEDURES

In this experiment, a population of 71 elementary level French learners at UVM in the Fall 2023 was selected for participation. An 'easier'/simple grammar structure (the near future) and a 'harder'/complex one (partitive articles) were selected, both of which should constitute novel grammar lessons to the students in elementary French. These choices parallel the criteria for selection of the model structures used by Yalçin & Spada (2016), but with students with less experience in their target L2 and by employing different tests to investigate the connection between cognitive function and FLA of the participants. The procedure included administration of aptitude tests (the LLAMA battery), a WM test (Automated Symmetry Span Task), and proficiency tests as measured by French Grammaticality Judgment Tasks (GJTs). The data collected were analyzed via Spearman's correlation coefficients to recover statistically significant relationships between variables, with the intention of identifying predictors of success at learning the target grammatical structures in French. It must be noted that while French 1100 at UVM was intentionally selected as a beginner level course, students did have varying levels of past exposure to the language as suggested by scores of a French placement test taken as a prerequisite to enrollment in the Fall 2023 semester (see Appendix D). Among 23 participants, scores ranged from 0 to 356, on a scale which recommends scores above 226 for enrollment in Elementary French II (a more advanced course than the one targeted in this study). Thus, despite the goal of achieving a participant pool of true beginners, it may be the case that one or both structures were already familiar to some of the participants tested for our study.

3.1 Target Structures

Two target structures in French were selected for this study that were expected to have varying levels of difficulty for learners. The participant pool was exposed to the near future (or

futur proche) grammatical construction which is quite similar to that of English. This pattern involves conjugating the verb *aller*, meaning *to go*, followed immediately by the infinitive form of the main verb (as seen in (1) below). The construction is nearly identical to the English equivalent, beside the fact that French employs this tense in the specific context of actions or happenings to occur relatively soon, turning instead to the future simple tense for events further ahead in time (Amon et al., 2019). In English, this "*going to*" construction can be more easily interchanged with the simple future tense. Nonetheless, the structure was hypothesized to be easier for new learners to acquire because of a lack of crosslinguistic obstacles.

1) *I / am going / to walk. = Je / vais / marcher.*

   Subject / conjugated aux verb / infinitive

In the case of the partitive articles, the complexity of the structure contributes to the purported difficulty of acquisition. While English is a language in which articles and/or quantifiers can often be omitted in the grammar, particularly for indefinite objects, this is not the case for French. All nouns (aside from pronouns) require an attached article or quantifier. Partitive articles exist as essentially the French equivalent of "some" or "any" in English, used to signify a non-count noun or a portion of a larger collection of items. For example, "some water" or simply "water" are acceptable noun phrases, but "a water" or "one water" are not. The four partitive articles are shown in Table 1 below:

Table 1. French Partitive Articles

	Number	**Gender**	**Contraction**
du	Singular/non-count	M	de + le
de la	Singular/non-count	F	——
de l'	Singular/non-count	M or F	de + le/la →vowel
des	Plural	M or F	de + les

As shown in Table 1, knowing when to use partitive articles and the ability to select the appropriate form involves several cognitive and grammatical operations to occur in quick succession. Conceptually, a speaker must not only have a grasp of which nouns are "non-count", but quickly make a series of decisions regarding the appropriate choice of article based on number, gender, phonetic agreement (preceding vowels vs. consonants), and in some cases, contraction. For native speakers, conscious effort is not required for use of these forms. However, the number of grammatical operations required due to the crosslinguistic differences explained before were hypothesized to render this structure more difficult for L2 learners of French. Despite their frequency in the input, they are not constrained to a particular context or situation like the near future conjugation, and thus may be seen as more variable and thus harder to use with high accuracy for new elementary level learners until this skill can become automatized (DeKeyser, 2000).

3.2 Grammaticality Judgment Tasks

The GJT used in this procedure was modeled after Yalçin & Spada (2016) but was necessarily simplified for effective implementation in an intact university class setting. Two written GJTs took the form of a list of pseudorandomized sentences in the L2, and an accompanying answer sheet on which participants marked each sentence either as grammatically acceptable or unacceptable. Differently from Y&S's procedure, participants in the context of this study were only given two options ("I don't know" was not a selection) and were not asked to make corrections to ungrammatical sentences. With the intent of respecting the intact classroom setting, instruction was not altered from the standard syllabus and lesson plans by the researcher at any level and remained entirely up to the professor's materials and schedule in the three sections of elementary French targeted in this study.

Since the GJTs were administered during normally scheduled class periods, in the interest of conserving time to minimize class interruptions, one combined pretest was favored over separate modules. The pretest with 60 total items, including 20 items per structure (near future and partitives), along with 20 additional filler items, was administered in the sixth week of the semester prior to the beginning of instruction for either target structure. In week eight, the near future was introduced. The posttest for this structure was administered the following week, nine days (four lessons) after initial exposure to the new material. The same timeline followed for the partitive articles, with the material introduced in the tenth week of the semester and the posttest occurring in week eleven. Though split into individual tasks, each posttest was comprised of the same 20 target structure items and 20 filler items that were shown on the pretest, in a differing order (see Appendix A for a sample of these materials). Each test was timed, allowing a baseline of thirty seconds per item (the pretest lasted thirty minutes; the posttests lasted twenty). However, it must be noted that while testing took place, all students were able to complete the full activity within the allotted time, often completing the task in less time than allotted.

3.3 Testing of aptitude & cognition

Beginning shortly after introduction of the project, eligible participants were able to utilize an online booking system to register for a one-time appointment in which the computer-based FLA and cognitive testing was administered individually by the researcher. These included LLAMA F, LLAMA B, LLAMA D, and LLAMA E, in the listed order. Followed by a short break, cognitive tests for attention (3-squared tasks), multitasking (SynWin), and working memory (Symmetry Span) were administered, only the last of which was included in the data represented within thesis project under the scope of the current study. The Automated Symmetry Span Task (Draheim et al., 2018) is a completely language-independent assessment of WM

capacity which has been widely used by linguistic and cognitive psychology researchers in the past. This test employs a visuospatial approach to assess working memory processes separately from language skills. Participants are shown a series of images which they must evaluate based on symmetry (reflecting the processing aspect of WM), while simultaneously being asked to memorize and then reproduce a separate sequence of shaded blocks on a grid, calling on the storage component of WM (Baddeley, 2003). Appendix C includes images from the Symmetry Span task for reference. The data collected from this task was divided into two scores: the absolute score measures the number of *sets* of items the participant recalled correctly, while the partial score totals the number of correct items cumulatively. Following completion of data collection, participant data were entered into non-parametric Spearman's correlations for analysis.

4. RESULTS

Table 2 (below) contains descriptive values for the LLAMA battery and WM test conducted, as well as for both the GJT pretest and posttest items for both structures. Means and standard deviations for each variable are included after having excluded 3 participants due to incomplete datasets and one for appearing as an extreme outlier for the partitive posttest, arriving at a total of n=19 datasets (out of the 71 participants tested for this project).

Table 2. Descriptive data for the LLAMA battery, WM task, & GJTs

DESCRIPTIVE STATISTICS	MEAN	STD. DEVIATION
n = 19		
LLAMA Avg	9.1447	2.38239
LLAMA F	8.3654	2.96668
LLAMA D	9.5789	3.00584
LLAMA E	7.3158	4.42283
LLAMA B	11.3158	4.47279
WM Abs.	17.1053	6.00828
WM Part.	27.5263	5.74812
NF Pretest	11.6842	3.35083
NF Posttest	15.5789	1.64370
Partitives Pretest	9.6842	2.21241
Partitives Posttest	10.1053	1.66315

Secondly, a paired samples T-test analyzing the pre- and post-GJTs for each structure was conducted, with results shown in Table 3. For the 18 participants who completed both parts of the near future GJT, the difference between the pre- and the posttests was statistically significant (t =-5.201, (18), p< .001), a result that indicates the change in scores after instruction was not due to chance.

Table 3. Paired Samples T-Test for GJTs

	t	df	One-sided P	Two-sided P
PreNF - PostNF	- 5.201	18	< .001	< .001
PrePart – PostPart	- .508	17	.309	.618

However, for the 17 participants who completed both sections for the partitive GJTs, the same did not obtain: the difference between pre and post test scores for the Part (partitive) structure did not achieve statistical significance. These results show that, while the participant data demonstrated significant improvement for the “easy” NF structure, the same was not true in the case of the partitives. As the partitive GJTs did not yield significant results, the following

analysis will focus mainly on the near future structure and the resulting change in proficiency detected in our study.

Table 4. Spearman correlations for LLAMA battery, 3-squared WM task, & GJT

	LL_AVG	LL_F	LL_D	LL_E	LL_B	WM_ABS	WM_PART	POSTNF	POSTPART
LL_AVG	-	.619**	.656**	.705**	.560**	.491*	.384	.156	-.316
LL_F		-	.508*	.487*	-.155	.125	.275	.101	-.216
LL_D			-	.272	.259	.411*	.476*	.248	-.439*
LL_E				-	.156	.074	-.086	-.120	.142
LL_B					-	.458*	.292	.227	-.216
WM_ABS						-	.739**	.339	-.058
WM_PART							-	.582**	-.330
POSTNF								-	-.081
POSTPART									-

*p < 0.05 **p < 0.01

Table 4 shows Spearman correlation coefficients for the four LLAMA battery components, WM assessment, and the GJT posttest results for both the easy and difficult structures. A final number of 19 participants were included in these correlations, after removal of extreme outliers as well as an additional three participants who were not able to complete all necessary components of the GJT pre- and post-tests due to class absences and/or scheduling conflicts. As would be expected, the LLAMA average scores highly correlated with each subtest measure. LLAMA F scores also correlated significantly with LLAMA D ($r\ (19) = .508$, $p < 0.05$) and with LLAMA E ($r\ (19) = .487$, $p < 0.05$). More importantly to the central aim of this study were the results for LLAMA D, assumed to be a measure of implicit learning (Granena, 2013), which correlated with both the scores for absolute WM score ($r(19) = .411$, $p < 0.05$) and partial WM score ($r(19) = .476$, $p < 0.05$). So did LLAMA B, though only with the absolute score ($r\ (19) = .458$, $p < 0.05$). The only variable found to have statistical significance in terms of predictive capacity for learning was the partial score for WM and was in fact highly correlated ($r\ (19) = .582$, $p < 0.01$) with PostNF, the post GJT for the French near future (easy) structure.

Given preceding description of results, a large portion of the following discussion will be devoted to the explanation of the relationship between WM capacity and its interaction with the "easy" structure, the near future, in our dataset. It will, however, be supplemented by the discussion of additional data points which approached statistical significance, as these may hold crucial implications for future research conducted with larger participant pools.

5. DISCUSSION

We will structure this discussion based on the original research questions posed earlier. We reiterate those questions below for the readers' sake.

R1) *Can FLA scores as measured by LLAMA v.3 successfully predict L2 proficiency of students in an elementary level French course?*

We predicted that there would be a positive correlation between the LLAMA tests and success at learning, particularly, the "difficult" structure in this experimental setup. However, our FLA scores did not predict proficiency on the GJT (Table 4). There are several possible reasons as to why this was the case. First, it may be the case that WM is less important to classroom settings because they tend to place less emphasis on real-time processing in the L2 as opposed to naturalistic settings (Foryś-Nogala,2021), which are infrequently studied in relation to FLA. The LLAMA battery is by no means a perfect tool for assessment of FLA (Bokander & Bylund, 2020), as demonstrated in Yalçin & Spada's (2016) finding that LLAMA scores showed no significant correlation with their spoken OPT (Oral Proficiency Task) scores despite their prediction. It may be that further reconsideration must be done to increase the validity of the subtests, or that the LLAMA scores sometimes correlate with proficiency measures due to other domain-general cognitive abilities which are involved in processing during these language-specific tasks (Wen, 2016, 2019).

R2) *Do students rely more heavily on their FLA capacity when tasked with learning a more challenging structure, as opposed to a simpler one?*

As mentioned in the results section, the data collected from the partitive GJT showed that no significant learning had occurred (Table 3). There are several factors which may have influenced the participants' capacity to learn the structure, considering the hypothesized relative grammatical difficulty of the structure in comparison to the near future. As was emphasized in the preceding description of the target structures, participants may still have been in the process of understanding the concept of partitives, both in terms of when to employ them conceptually and how to do so accurately. Though both structures were predicted to tap explicit knowledge resources, the number of steps (conceptual recognition of non-count nouns, gender/plural agreement, contractions, vowel/consonant agreement) required to properly use partitive articles may have been lacking in our participants' performance during the testing phases. Not only did the two target structures selected for this research vary in terms of complexity, but the near future conjugation involves usage of content words central to the overall meaning of the utterances provided. Consequently, if our participants wanted to derive meaning from those sentences, they necessarily had to process the full grammatical construct to derive a semantic representation. In the case of the partitives, conversely, the relevant structure consists only of function words and is included in a much broader variety of sentences, applicable to a much larger range of topics. It may very well be the case that the near future is a much more salient form to the participants, as this verbal construction constitutes a critical piece in comprehension of the utterance. Because the accurate use of a partitive article does not encode crucial information to the overall meaning of the sentence, our participants may have disregarded it and paid more attention to the lexical items instead of those functional forms.

Some additional obstacles to learning may have occurred simply due to the timeline and format of the project as implemented in an intact classroom. Completion of this research project was dependent upon the existing timeframe of the university's semester schedule and the professor's required materials to cover within this same period. As such, both the near future and partitive posttests were administered nine days after initial introduction, and thus the partitive GJT may have taken place too soon relative to the difficulty of the structure. Furthermore, it must be noted that participants were incentivized to take part in the study in the form of extra credit points allocated to their final grade in the course. Approximately one third of the eligible participants chose to self-enroll in the aptitude and cognitive testing. It could then be surmised that our dataset included a higher proportion of students who perceived their language skills to be below expectations or had low motivation to acquire the L2 successfully.

Significant learning was demonstrated by the difference in our results between the pre- and posttests for the near future GJT. In contrast to the partitive structure, success at learning this construction may be due to the simpler, formulaic process needed to form the near future. As long as participants were able to recognize and recall the *subject + aller (conjugated) + infinitive* sequence, which they were already familiar with, they could rely on that guideline and correctly identify correct vs incorrect versions of the GJT forms. Remember that participants were not asked to produce or edit forms in any way, just to judge their grammaticality.

R3) *What is the relationship of working memory with FLA and with proficiency?*

As we expected, WM does have significant influence on proficiency in the L2 in the context of SLA (Wen, 2021;2023). In our dataset, the particular relationship highlighted in our findings was that partial credit WM scores were shown to predict approximately 38% of variance in the near future GJT results (Table 4). Returning to the Aptitude Profile Model as outlined by

Skehan (2002), it may be inferred that WM is relevant to learning during the patterning stage in SLA, as participants in elementary level French begin to navigate recognition and usage of conjugations in the near future tense. Since learning did not take place for the difficult target structure, the lack of a correlation between WM and the GJT for the partitives may be attributed to the absence of improvement in the noticing and patterning developmental stages in Skehan's framework.

Our data presented additional significant correlations between some of the LLAMA subtests and WM scores. This finding is in partial support of the hypotheses behind R3, particularly when highlighting that the absolute WM score significantly correlated with the mean scores for the LLAMA battery, and WM capacity explained 24% of the variance on the average scores for all LLAMA tests. Both, the absolute and the partial working memory scores predicted performance in the LLAMA D sub-task, perhaps due to the implicit nature of the task (Iizuka & DeKeyser, 2023) which prompts short-term storage of auditory input and fast-paced processing required to correctly identify test items. LLAMA B, a test devoted to vocabulary learning, also relies heavily on memorization of items during the testing phase and the absolute WM score accounted for approximately 21% of variance for this test. Previous studies have shown again and again that some robust connection exists between language use and WM (Li, 2016), the latter capacity having been correlated with a diverse set of applied tasks and functions due to its overlapping role in storage and processing of information.

What is more, the positive correlations of LLAMA F with both LLAMA D and LLAMA E scores suggest that this FLA test battery does effectively measure some core component of linguistic ability needed to succeed in the acquisition of an L2. In spite of the small number of participants tested in our study, the results do indicate that aspects of FLA are positively

correlated to WM, which, in turn, does show a positive and significant correlation with improvement in the acquisition of a grammatical structure in French as an L2. Future studies which replicate the methodology we followed on a larger scale would likely reveal a more comprehensive set of correlations between WM, FLA, and proficiency in the L2.

6. CONCLUSIONS

Our top finding demonstrates that WM is partially responsible for the acquisition of new material in a second language classroom in a university setting. We also found significant positive correlations between multiple components of FLA as distinguished by the LLAMA subtests with WM, demonstrating cognitive overlap between these abilities when deployed in a language classroom context. Similarly to Yalçin & Spada's (2016) finding with their OPT data, we found that FLA measures did not correlate with GJT performance, and thus we were not able to support the hypothesis that the LLAMA test scores would predict performance on the GJT posttests administered to the elementary level French learners. However, the positive correlations identified in the preceding discussion acknowledge the relevance of WM to both FLA and actual proficiency measures (GJT). It is clear that the analysis produced in our study was significantly limited by the lack of learning shown by our participants with the difficult structure, which could be attributed to several factors. The main limitation when analyzing our results stems from the small size of our participant pool, in spite of our many efforts to entice all 71 students in elementary French to participate in our study. Nonetheless, this study highlights the need for further research on both FLA and WM in relation to SLA. This is especially true because the most current version of the LLAMA battery has still been underemployed in empirical research when compared to its earlier version, although the correlations shown between the LLAMA sub-tasks point to an underlying construct closely linked to linguistic ability in an L2.

Acknowledgments

I would like to first and foremost thank my thesis advisor, Dr. Guillermo Rodríguez, for his guidance, expertise, and encouragement throughout the entirety of this project and for welcoming me into the SLUVM lab. Your kindness and brilliance have left a mark on me in my time at the University of Vermont that will not be forgotten. I would also like to thank Dr. Ching Selao and Dr. Ben Sienicki, for choosing to sit on my committee, and for sharing in your intelligence and experience with me as a student as I found my place within the French and Linguistics departments. Thank you to Professor Thomas Dunn, for your willingness to allow my research to take place in your classroom and for supporting my work. More than anything, I would like to thank my family and friends for their enduring support throughout my youth and my college days: my grandparents Rick & Mary Simmons, my late father Corey Gamache, my dear roommates at Red House, and friends of Dude City & Reggie's house. To Katie, thank you for allowing me to experience the joy of being a sister in this life, for sharing so much childhood and laughter with me. And to Mom, thank you for being a beacon in my life, an image of a brilliant, driven, and accomplished woman, yet who has shown me endless love and solidarity no matter where my ambitions took me.

References

Ameringer, V., Green, L., Leisser, D., & Turker, S. (2018). Getting to the Core of Language Aptitude. In S. M. Reiterer (Ed.), *Exploring Language Aptitude: Views from psychology, the language sciences, and cognitive neuroscience.* (pp. 5-11). Springer International Publishing AG.

Amon, E., Muyskens, J., & Ommagio Hadley, A.C. (2019). *Vis-à-vis: Beginning French* (7th ed.). McGraw Hill Education.

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

Bell, S. M., & McCallum, R. S. (2012). Do Foreign Language Learning, Cognitive, and Affective Variables Differ as a Function of Exceptionality Status and Gender? *International Education, 42*(1), 86-106.

Bokander, L. (2020). Language Aptitude and Crosslinguistic Influence in Initial L2 Learning. *Journal of the European Second Language Association*, *4*(1). https://doi.org/10.22599/jesla.69

Bokander, L. and Bylund, E. (2020), Probing the Internal Validity of the LLAMA Language Aptitude Tests. Language Learning, 70: 11-47. https://doi.org/10.1111/lang.12368

Carroll, J. B. (1958). A factor analysis of two foreign language aptitude batteries. *The Journal of General Psychology, 59*(1), 3-19.

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

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

Carroll, J.B., & Sapon, S. (1959). *Modern language aptitude test.* Bethesda, MD: Second Language Testing.

DeKeyser, R. M. (2000). The robustness of critical period effects in second language acquisition. *Studies in Second Language Acquisition, 22*(4), 499–533.

Draheim, C., Harrison, T. L., Embretson, S. E., & Engle, R. W. (2018). What item response theory can tell us about the complex span tasks. *Psychological assessment*, *30*(1), 116–129. https://doi.org/10.1037/pas0000444

Engle, R. W. (2018). Working Memory and Executive Attention: A Revisit. Perspectives on Psychological Science, 13(2), 190-193. https://doi.org/10.1177/1745691617720478

Foryś-Nogala, M. (2021). Cognitive Aptitudes and Processing of L2 Grammar: Exploring the Role of Rule Inferencing Skills and Working Memory. *Advances in Cognitive Psychology*, *17*(4), 310-319. https://doi.org/10.5709/acp-0340-1

Granena, G. (2013). 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

Hyltenstam, K. (2021). Language aptitude and language awareness: Polyglot perspectives. *Annual Review of Applied Linguistics*, *41*, 55-75.

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

Li, S. (2016). The Construct Validity of Language Aptitude. *Studies in Second Language*

Acquisition, *38*(4), 801-842. https://doi.org/10.1017/s027226311500042x
Li, S., & Zhao, H. (2021). The methodology of the research on language aptitude: A systematic review. *Annual Review of Applied Linguistics, 41*, 25-54. doi:10.1017/S0267190520000136
Long, M.H. (2007). Problems in Second Language Acquisition (1st ed.). Routledge. https://doi.org/10.4324/9781315089447
McLaughlin, B. (1987). Theories of second-language learning. London: Edward Arnold.
Meara, P. (2005). *LLAMA Language Aptitude Tests.* Lognostics.
Meara, P. & Rogers, V. (2019). *The LLAMA Tests v3.* Cardiff: Lognostics.
Mujtaba, S. M., Kamyabi Gol, A., & Parkash, R. (2021). A study on the relationship between language aptitude, vocabulary size, working memory, and L2 writing accuracy. *Foreign Language Annals*, *54*(4), 1059-1081. https://doi.org/10.1111/flan.12584
Ortega, L. (2009). *Understanding second language acquisition*. Hodder Education.
Rogers, V., Meara, P., Barnett-Legh, T., Curry, C., & Davie, E. (2017). Examining the LLAMA aptitude tests. *Journal of the European Second Language Association*, *1*(1), 49-60. https://doi.org/10.22599/jesla.24
Rogers, V., Meara, P., & Rogers, B. (2023). Testing Language Aptitude: LLAMA Evolution and Refinement. In Z. (Edward) Wen, P. Skehan, & R. L. Sparks (Eds.), *Language Aptitude Theory and Practice* (pp. 47–72). chapter, Cambridge: Cambridge University Press.
Saito, K., Suzukida, Y., & Sun, H. (2019). APTITUDE, EXPERIENCE, AND SECOND LANGUAGE PRONUNCIATION PROFICIENCY DEVELOPMENT IN CLASSROOM SETTINGS: A LONGITUDINAL STUDY. *Studies in Second Language Acquisition, 41*(1), 201-225. https://doi.org/10.1017/S0272263117000432
Skehan, P. (2002). Theorising and updating aptitude. In P. Robinson. (Ed.), Individual differences and instructed language learning (pp. 69 – 93). Amsterdam, The Netherlands: John Benjamins.
Skehan, P. (2015). Brief articles on Aptitude and Attention from the Routledge Encyclopedia of SLA
Sparks, R. L., Patton, J., Ganschow, L., Humbach, N., & Javorsky, J. (2006). Native Language Predictors of Foreign Language Proficiency and Foreign Language Aptitude. *Annals of Dyslexia*, *56*(1), 129–160. https://doi-org.ezproxy.uvm.edu/10.1007/s11881-006-0006-2
Sparks, R. L., Artzer, M., Ganschow, L., Siebenhar, D., Plageman, M., & Patton, J. (1998). Differences in Native-Language Skills, Foreign-Language Aptitude, and Foreign-Language Grades among High-, Average-, and Low-Proficiency Foreign-Language Learners: Two Studies. *Language Testing, 15*(2), 181-216. DOI:10.1191/026553298669285653
Wen, Z., & Li, S. (2019). Working Memory in L2 Learning and Processing. In *The Cambridge Handbook of Language Learning* (pp. 365-389). https://doi.org/10.1017/9781108333603.016
Wen, Z. E. (2021). Language Aptitudes. In T. Gregersen & S. Mercer (Eds.) The ‘
Wen, Z. (Edward), Skehan, P., & Sparks, R. L. (Eds.). (2023). *Language Aptitude Theory and Practice*. Cambridge: Cambridge University Press. Routledge Handbook of the Psychology of Language Learning and Teaching
Yalçın, Ş., & Spada, N. (2016). Language aptitude and grammatical difficulty: An EFL classroom-based study. *Studies in Second Language Acquisition*, *38*(2), 239-263.

APPENDIX A
FRENCH GRAMMATICALITY JUDGMENT TASK: Proficiency Task Sample Items

Directions: Read each sentence through and answer on your scantron sheet whether each is an acceptable (A) or *un*acceptable (B) sentence in French. There are no errors based on spelling (focus on grammar, agreement, conjugations, etc.) If you are not sure, fill in a guess. You will have 30 minutes to complete the activity.

1. Elle va choisir une majeure à l'université.
2. Je prends du café, s'il vous plaît.
3. Qui vont choisir la musique ?*
4. Je parle un peu du français.*
5. Nous allons chercher nos amis.
6. Mes amis et moi jouons du la musique.*
7. Où allez-vous partir en vacances ?
8. Elle va une petite sœur.*
9. Nous avons chaud pendant l'été.
10. Les mères sont gentilles et intelligentes.
11. Tu ne vas pas aller à l'école aujourd'hui.
12. J'ai un grand sœur et un petit frère.*
13. Maria vient d'Espagne.
14. Les étudiants se préparent pour la examen.*
15. Tu vas aller au parc demain ?

***ungrammatical item**

APPENDIX B
LLAMA v.3: Foreign Language Aptitude

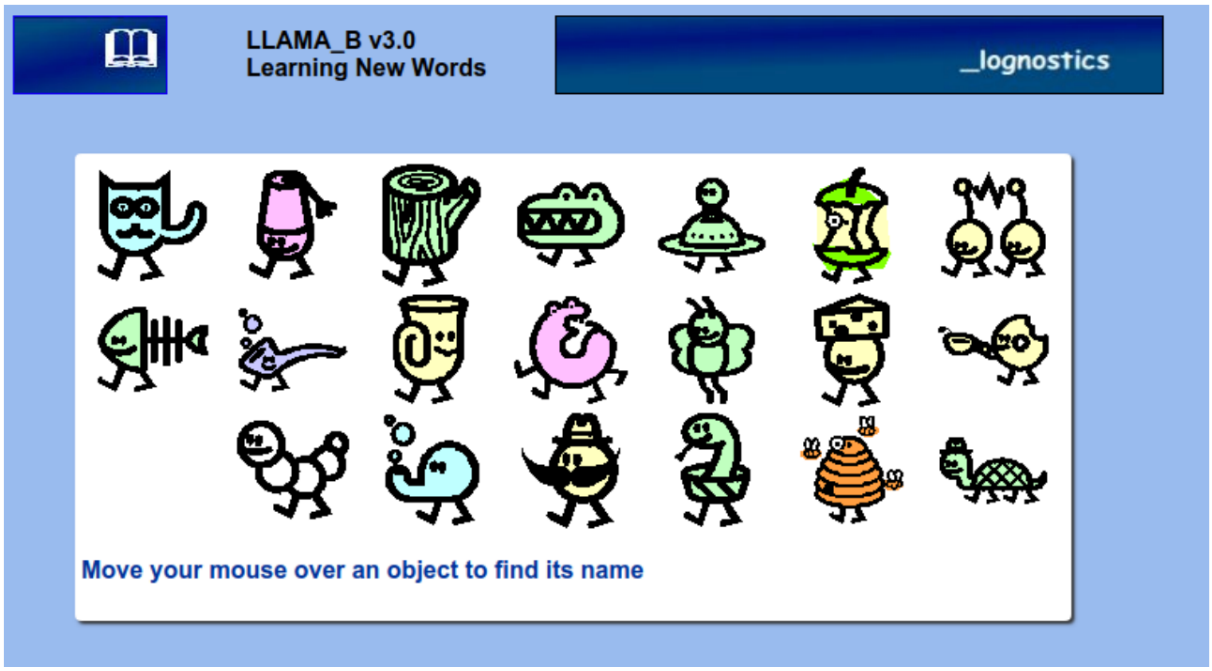


Figure 1. LLAMA B learning phase

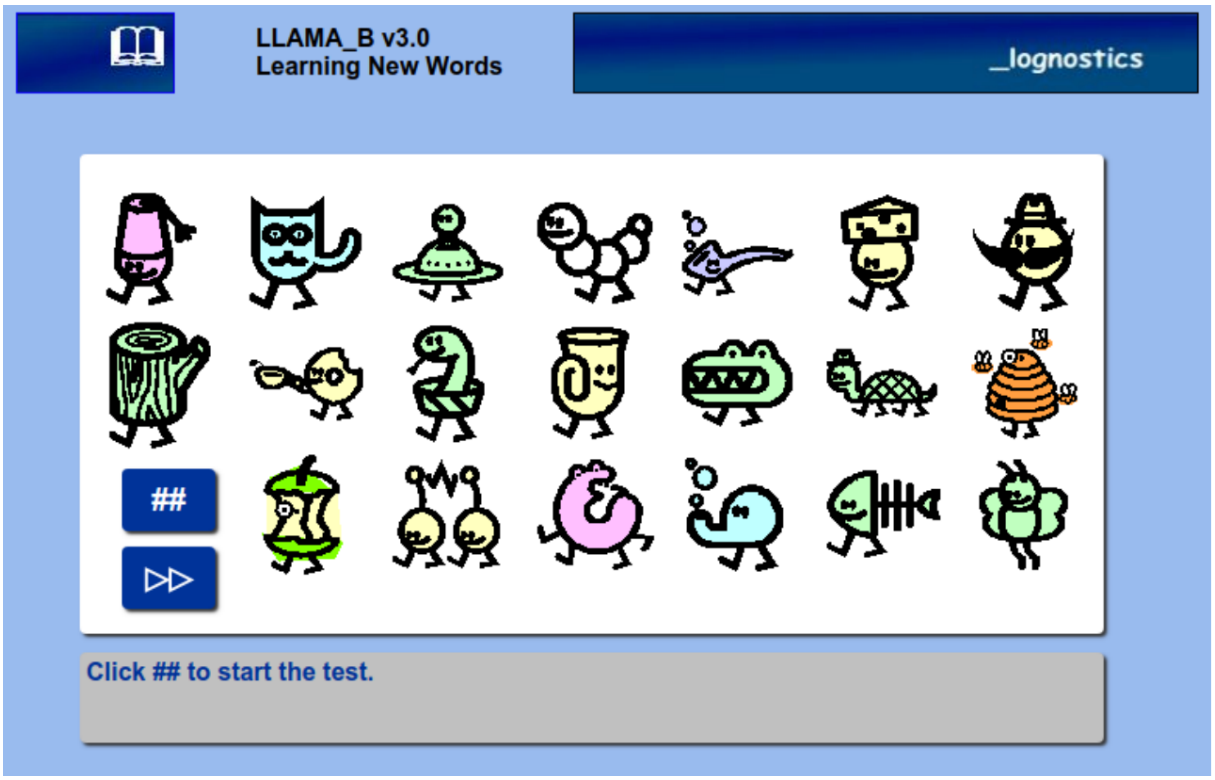


Figure 2. LLAMA B test page

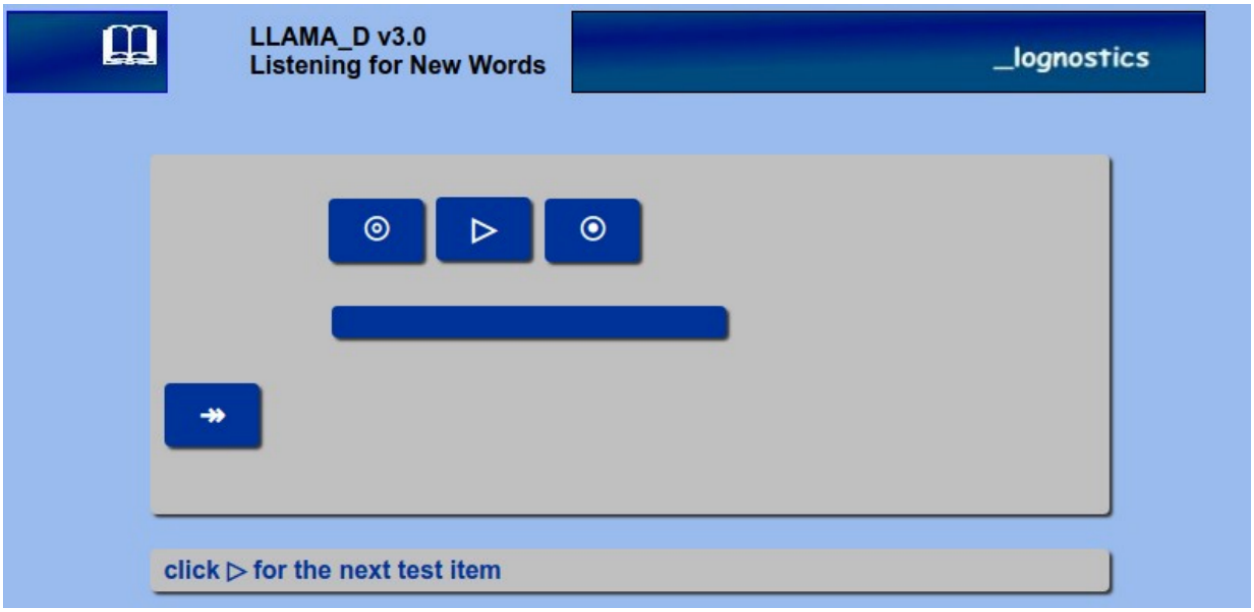


Figure 3. LLAMA D test page

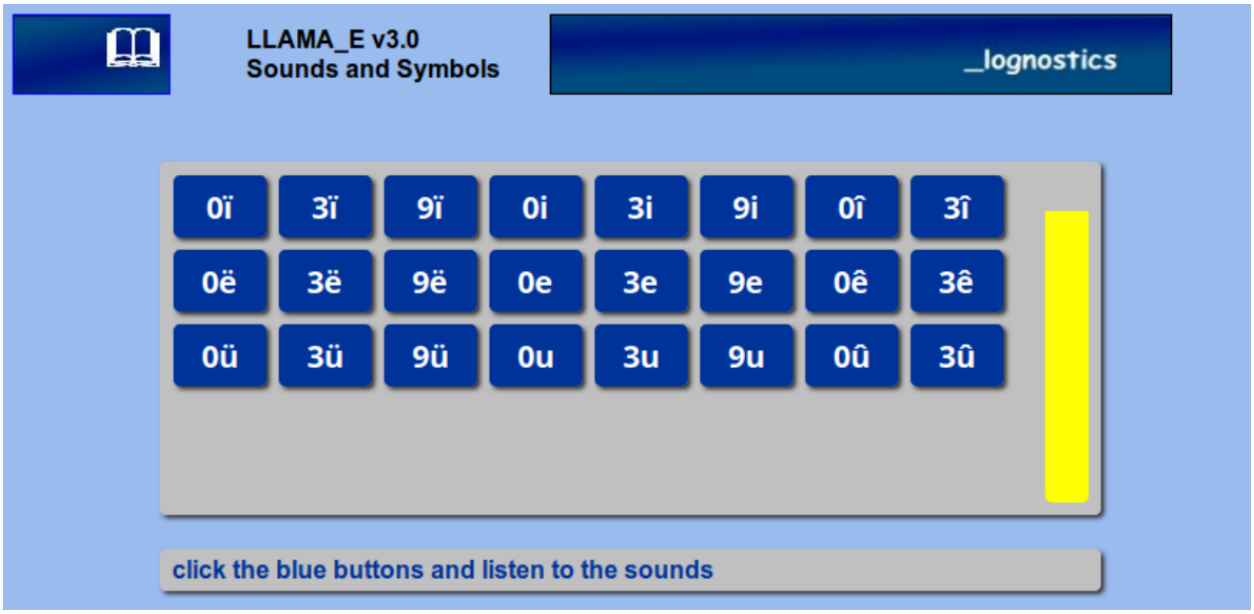


Figure 4. LLAMA E learning page

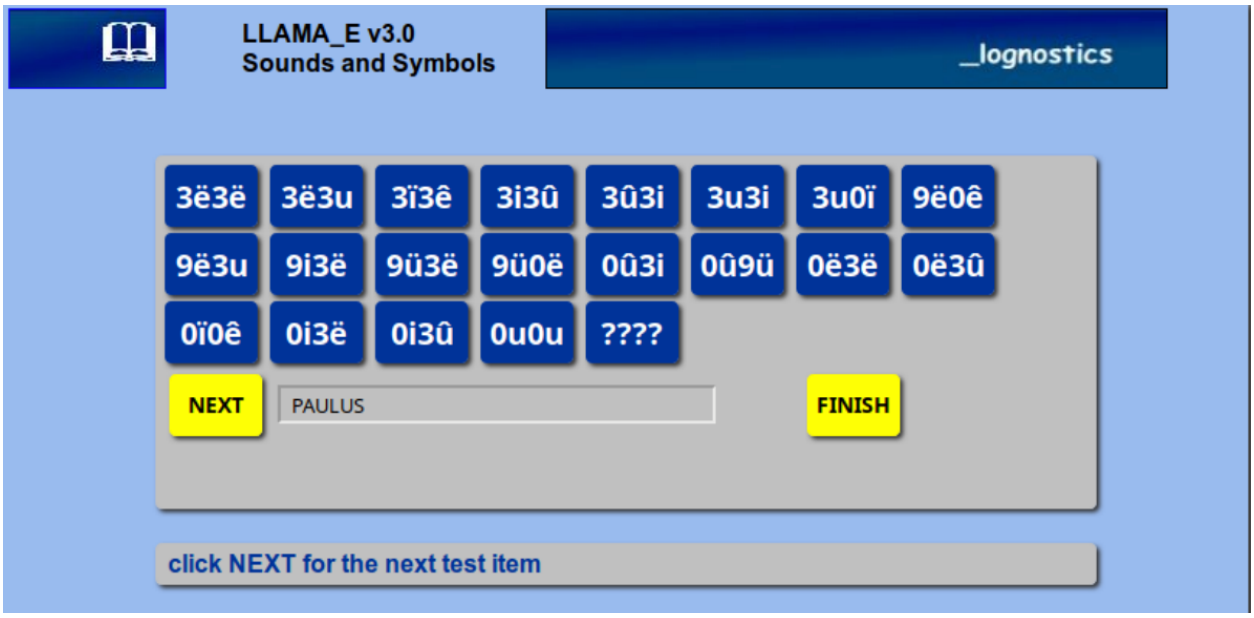


Figure 5. LLAMA E test page

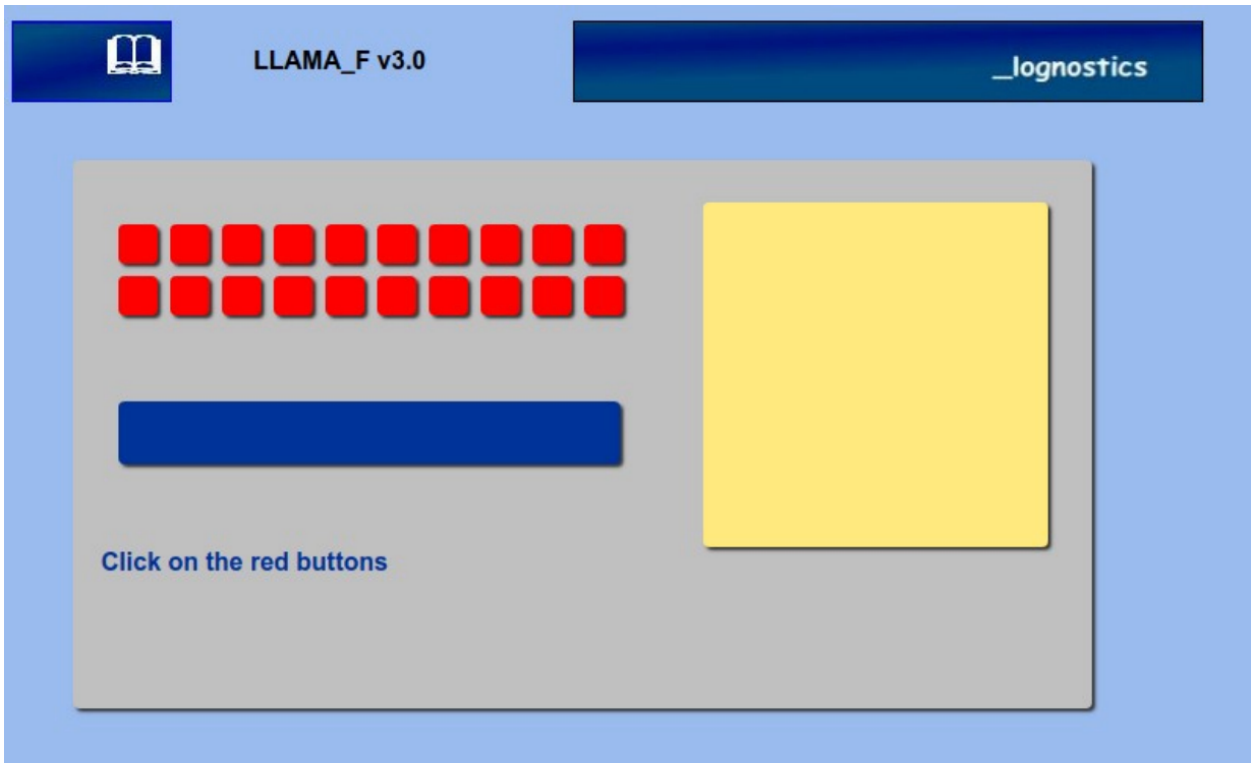


Figure 6. LLAMA F test page

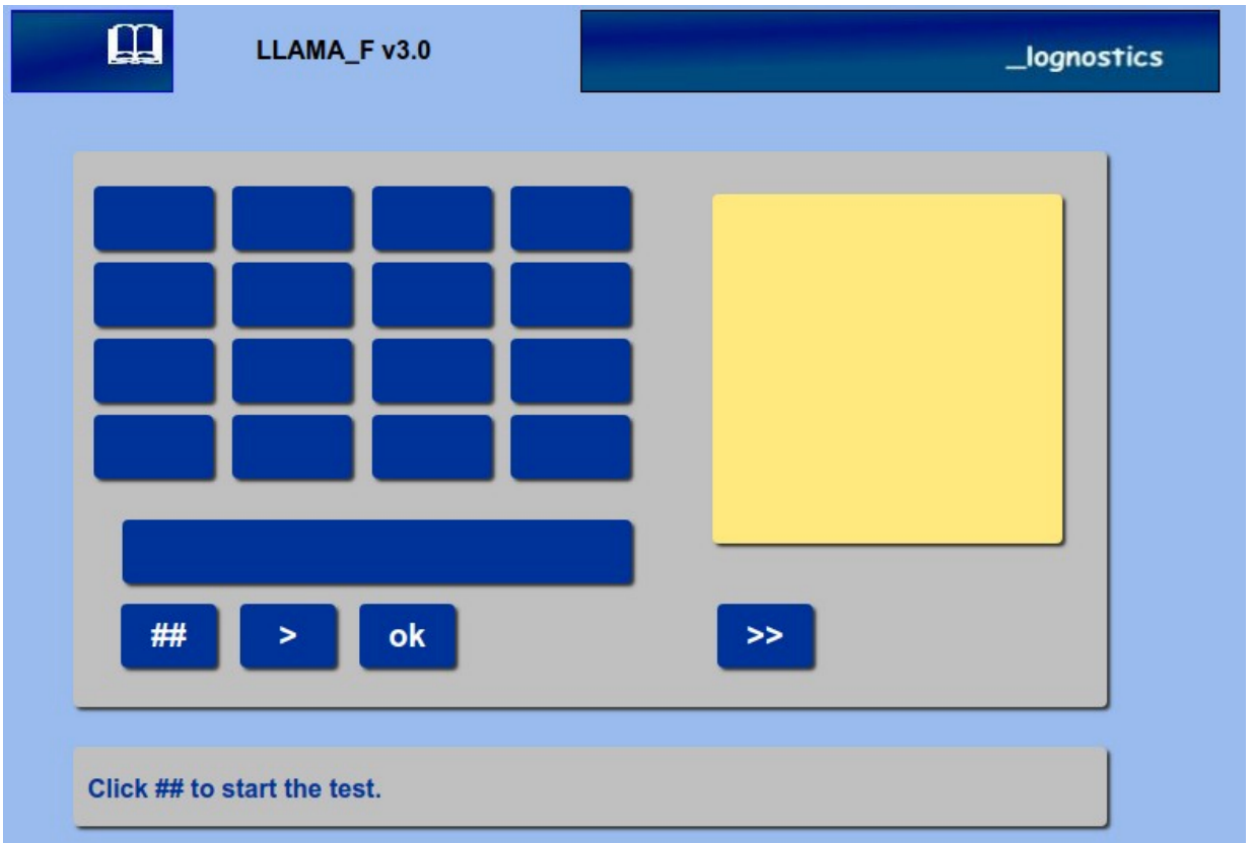
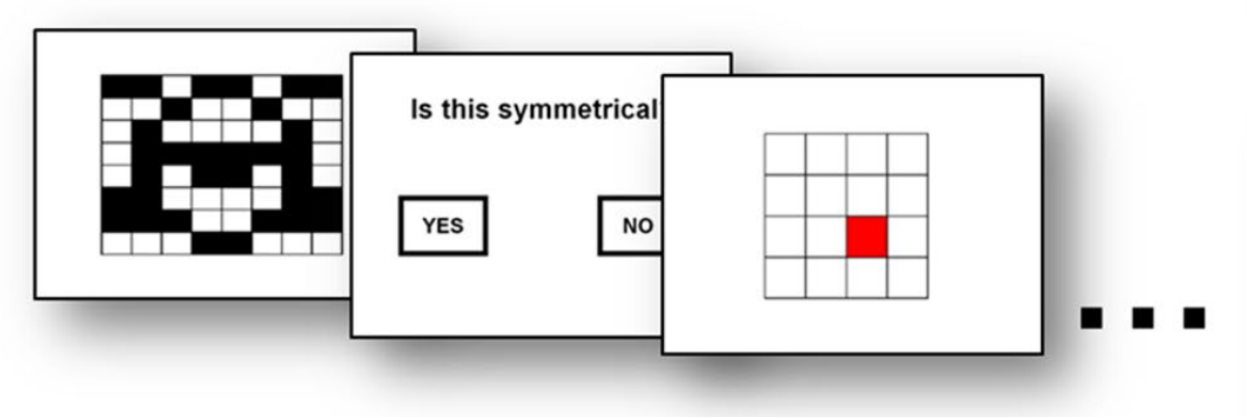


Figure 7. LLAMA F test page

APPENDIX C
AUTOMATED SYMMETRY SPAN: Working Memory Capacity

Figure 8 shows the trial structure for one item of a set in this task.



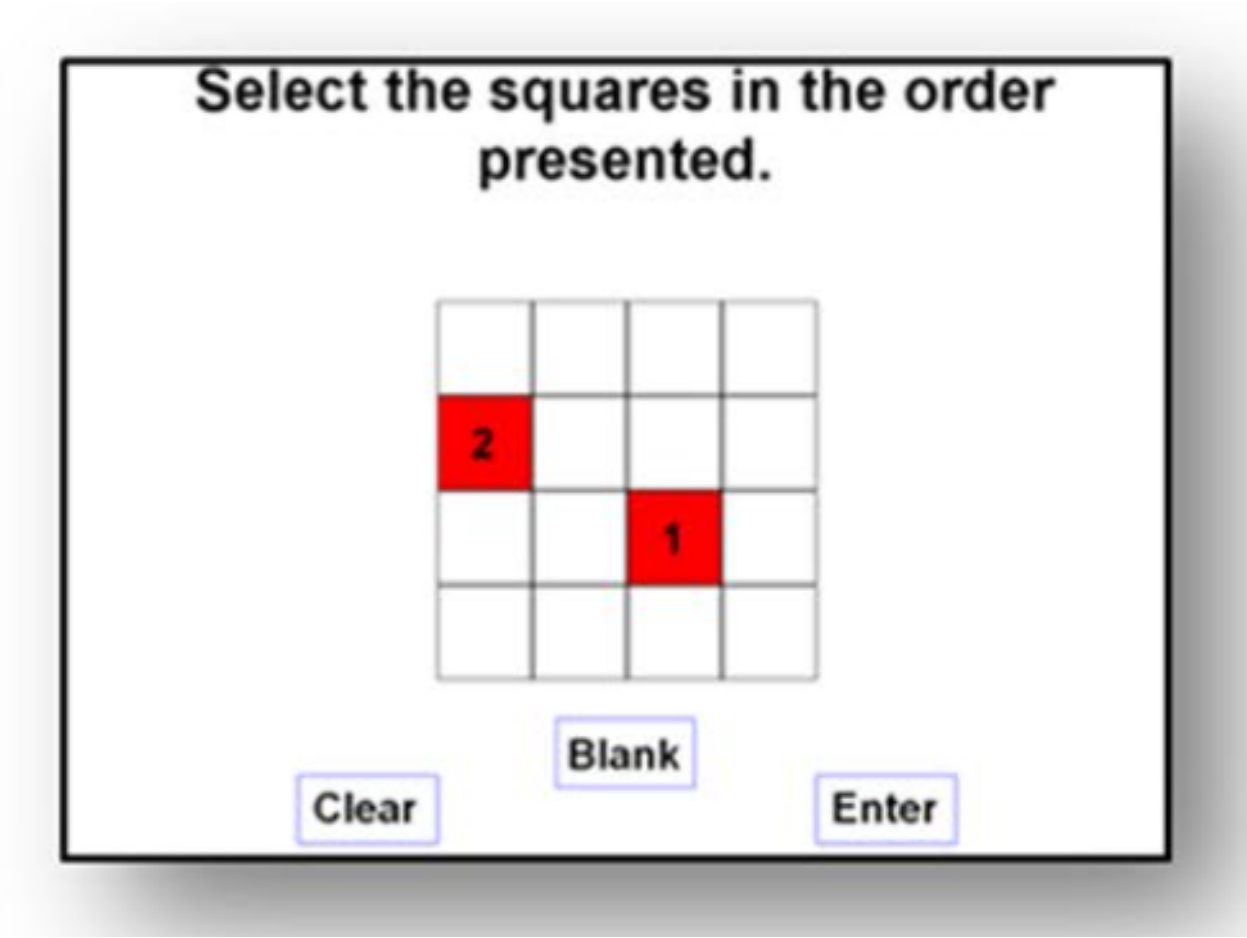


Figure 8. Experimental trial sequence for the Automated Symmetry Span task.

APPENDIX D
FRENCH PLACEMENT TEST SCORES

Table 6. French placement scores for complete study participants

Participant ID	Placement score
23020001	23
23020002	217
23020003	0
23020004	145
23020005	137
23020006	0
23020007	0
23020008	0
23020009	0
23020010	141
23020011	125
23020012	153
23020013	0
23020014	—
23020015	0
23020016	25
23020017	0
23020018	356
23020019	173
23020020	72
23020021	100
23020022	0
23020023	0

Score	Recommended Level
Below 226	FREN 1100
226-300	FREN 1200
301-370	FREN 2100
371-440	FREN 2200
Above 440	FREN 3000 level*

Figure 9. Course level recommendations as per UVM website