



Universiteit  
Leiden  
The Netherlands

## Language neutrality of the LLAMA test explored: the case of agglutinative languages and multiple writing systems

Mikawa, M.; Jong, N.H. de

### Citation

Mikawa, M., & Jong, N. H. de. (2021). Language neutrality of the LLAMA test explored: the case of agglutinative languages and multiple writing systems. *Journal Of The European Second Language Association*, 5(1), 87-100. doi:10.22599/jesla.71

Version: Publisher's Version

License: [Creative Commons CC BY 4.0 license](https://creativecommons.org/licenses/by/4.0/)

Downloaded from: <https://hdl.handle.net/1887/3274273>

**Note:** To cite this publication please use the final published version (if applicable).



# Language neutrality of the LLAMA test explored: The case of agglutinative languages and multiple writing systems

MOMO MIKAWA

NIVJA H. DE JONG

*\*Author affiliations can be found in the back matter of this article*

RESEARCH

WHITE ROSE  
UNIVERSITY PRESS  
Universities of Leeds, Sheffield & York

## ABSTRACT

The ability to learn a foreign language, language aptitude, is known to differ between individuals. To better understand second-language learning, language aptitude tests, tapping into the different components of second-language learning aptitude, are widely used. For valid conclusions on comparisons of learners with different language backgrounds, it is crucial that such tests be language neutral. Several studies have investigated the language neutrality of the freely available LLAMA tests (Granena, 2013; Rogers et al., 2016, 2017). So far, comparing a number of L1 backgrounds, including those using different writing systems such as Arabic and Mandarin, no significant differences between participants have been found. However, until now, neither participants with agglutinative language backgrounds nor with first-language backgrounds that use multiple writing systems have been included. Therefore, this study selected participants from three different first-language backgrounds: Dutch (non-agglutinative, phonogram/Latin alphabet), Hungarian (agglutinative, phonogram/Latin alphabet), and Japanese (agglutinative, phonogram/syllabic alphabet and logogram/Japanese kanji). The participants performed three subsets of the LLAMA test. Significant differences between the groups were found on two of these tests: The ability to implicitly recognize sounds (LLAMA\_D subtest) and inductive grammar learning ability (LLAMA\_F), but no differences were found on vocabulary learning ability (LLAMA\_B). Additionally, for LLAMA\_B, the number of languages learnt was a significant covariate, confirming earlier findings that some subtests seem to be linked to language learning experience. We discuss the implications of our findings on the validity of the LLAMA\_D and LLAMA\_F subtests.

CORRESPONDING AUTHOR:

**Nivja H. de Jong**

Leiden University Center for  
Linguistics, Leiden University,  
NL; ICLON Graduate School of  
Teaching, Leiden University, NL  
[n.h.de.jong@hum.leidenuniv.nl](mailto:n.h.de.jong@hum.leidenuniv.nl)

KEYWORDS:

Aptitude; second-language  
learning; testing; validity;  
agglutinative languages;  
writing systems

TO CITE THIS ARTICLE:

Mikawa, M., & De Jong, N. H.  
(2021). Language neutrality  
of the LLAMA test explored:  
The case of agglutinative  
languages and multiple  
writing systems. *Journal of the  
European Second Language  
Association*, 5(1), 87–100.  
DOI: [https://doi.org/10.22599/  
jesla.71](https://doi.org/10.22599/jesla.71)

## 1. INTRODUCTION

Language aptitude has been regarded as a specific talent for learning foreign languages (Carroll, 1981; Skehan, 2002) and is known to vary considerably among individuals (Dörnyei & Skehan, 2003). Language aptitude is generally seen as consisting of several cognitive components, distinct from verbal intelligence and motivation (e.g., Carroll, 1973, 1981). Throughout the years, a number of language aptitude tests have been developed, such as the Modern Language Aptitude Test (MLAT; Carroll & Sapon, 1959), Pimsleur Aptitude Battery (Pimsleur, 1966), Defense Language Aptitude Battery (DLAB; Petersen & Al-Haik, 1976), Test of Cognitive Ability for Novelty in Acquisition of Language – Foreign (CANAL-FT; Grigorenko et al., 2000), Swansea Language Aptitude Tests (LAT; Meara et al., 2002), LLAMA (Meara, 2005) and High-Level Language Aptitude Battery (HiLAB; Linck et al., 2013). The different aptitude tests also reflect the somewhat different views on the construct itself (for an overview, see Li, 2016). In recent years, the LLAMA test has been widely used to measure language aptitude of second-language (L2) learners. A number of studies have investigated the validity of the LLAMA test (Granena, 2013; Rogers et al., 2016). These studies have concluded the test to be valid with respect to participants' language background. However, to date, agglutinative languages have not been taken into account as a first language (L1) nor have researchers investigated any language that makes use of a different writing system. Therefore, the present study addresses these language characteristics to explore the validity of the LLAMA test with respect to L1 background by comparing Hungarian (agglutinative language with phonogram writing system), Japanese (agglutinative language with phonogram and logogram writing systems) and Dutch (non-agglutinative language with phonogram writing system).

## 2. BACKGROUND

### 2.1. LANGUAGE APTITUDE IN THE LLAMA TEST

Language aptitude is seen as one of the characteristics that may explain individual differences in success in learning an L2 (Grigorenko et al., 2000). Language aptitude has been theorized to consist of a number of cognitive abilities (e.g., Carroll, 1973; Dörnyei, 2014). Phonemic coding ability, grammatical sensitivity, inductive language learning ability and associative memory are seen as key cognitive abilities for learning an L2 (Carroll, 1973), and these components are included in the MLAT developed by Carroll and Sapon (1959). Based on the MLAT, Meara et al. (2002) published the LAT tests that consist of five subtests, measuring the following elements: Phonetic memory skills, capacity of vocabulary learning, grammatical inferencing ability, memory ability for sequences of unknown sounds, and ability for sound symbol association. The LAT test was initially developed for L1 speakers of English, and it may be less suited for L1 speakers who use other orthographic systems than the Latin alphabet, such as Greek or Japanese. Moreover, one of the subtests of the LAT test is designed based on existing languages, which means the test cannot be used to measure aptitude of L1 speakers of these languages.

Based on the LAT test, Meara (2005) developed the LLAMA Language Aptitude Test, which is freely available (<http://www.lognostics.co.uk/tools/llama/>). The LLAMA test battery consists of four sub-components: The LLAMA\_B (vocabulary learning ability), LLAMA\_D (sound recognition), LLAMA\_E (sound-symbol association) and LLAMA\_F (grammatical inferencing). Recently (after data collection for this study had been completed), a new version of the LLAMA tests was published (Meara & Rogers, 2020).

### 2.2. PREVIOUS STUDIES

Language independence or language neutrality is a crucial characteristic for aptitude measures, meaning that the scores on the language aptitude test should not be affected by the writing systems, phonetic features, or grammar characteristics of the L1 of the test users. Especially if the LLAMA test is to be used in studies that measure language learning aptitude of L2 learners with diverse L1 language backgrounds, language independence is crucial. Likewise, when comparing across studies with diverse L1 backgrounds (e.g., when comparing Artieda & Munoz, 2016; Kepinska et al., 2017; Saito, 2017; Yalçın et al., 2016; Yılmaz & Granena, 2016), language independence of the test for aptitude is important. For this reason, several studies have investigated the validity and the reliability of the LLAMA test itself. These studies have focused on internal validity of the tests (Bokander & Bylund, 2020) and on a number of

individual variables of the participants, such as gender, age of onset (Granena & Long, 2013), multilingualism (Rogers et al., 2017), prior L2 instruction (Rogers et al., 2017) and L1 language background (Granena, 2013; Rogers et al., 2016, 2017). Here we briefly review the studies focusing on L1 language background.

Granena (2013) investigated the LLAMA test with 186 participants (79 males, 107 female) who were under the age of 40 ( $M = 25$ ,  $SD = 5.29$ ,  $range = 18-39$ ) with different language backgrounds: English, Spanish, and Chinese. The results showed that the scores did not differ depending on the L1. Furthermore, the results of her study drew the conclusion that the LLAMA test measured two different dimensions of language aptitude; the LLAMA\_B (vocabulary learning capacity), LLAMA\_E (sound-symbol association) and LLAMA\_F (grammatical inferencing) invoke explicit learning ability, whilst the LLAMA\_D (sound recognition) is more related to implicit learning ability. Moreover, in their study of 350 participants with diverse L1s (L1 groups with more than 20 participants were Afrikaans, English, German, and Wolof), Bokander and Bylund (2020) carried out a principal component analysis that also identified the subtest LLAMA\_D (sound recognition) as the odd one out. This finding supports the conclusion that LLAMA\_D, unlike the other tasks, taps into implicit processing.

Rogers et al. (2016) examined 229 participants ranging in age between 10 to 75 with diverse L1 backgrounds and educational levels. Because they wanted to compare participants who had learned different types of scripts in their L1, participants were divided into three groups: English Latin script ( $n = 99$ ), non-English Latin script ( $n = 18$ ), and non-Latin script (mostly Chinese;  $n = 17$ ).<sup>1</sup> They investigated whether differences in L1 writing systems impacted the scores of language aptitude. The results did not reveal any specific effects for the L1, but they did find an effect for educational level in three components: LLAMA\_B (vocabulary learning capacity), LLAMA\_E (sound-symbol association) and LLAMA\_F (grammatical inferencing).

Consequently, Rogers et al. (2017) carried out an even broader validation study of the LLAMA test. They examined the role of the L1 writing system and the language neutrality in general of the LLAMA test. They chose L1 speakers of English ( $n = 107$ ), Chinese ( $n = 56$ ) and Arabic ( $n = 32$ ) as participants. The L1 writing systems of the participants therefore consisted of the English Latin alphabet, Chinese morphosyllabic, Chinese logographic system (Tolchinsky et al., 2012), and the Arabic consonantal alphabet. The results showed that there were significant differences among these three groups in the LLAMA\_B, LLAMA\_E and LLAMA\_F, but not in the LLAMA\_D. It turned out, however, that these groups were not comparable because the number of monolinguals was different across groups. Therefore, they analysed a new dataset — L1 speakers of English ( $n = 48$ ), Chinese ( $n = 56$ ), and Arabic ( $n = 30$ ), excluding all monolingual participants. For these new groups, there were no significant differences. They concluded that with respect to these writing systems and language backgrounds, the LLAMA test is language neutral. Rogers et al. (2017) also followed up on research and hypotheses by Sparks et al. (1995), among others, that showed that language learning experience may have a positive influence on language learning aptitude scores. Among the variables investigated, they found that prior L2 instruction was the strongest predictor of aptitude, especially in LLAMA\_B (vocabulary learning ability) and LLAMA\_F (grammatical inferencing).

These studies have examined diverse L2 learners varying in L1 typology and different writing systems in order to investigate the validity of the LLAMA test, yet none of them has taken agglutinative languages into account as L1s, nor have they investigated languages with multiple writing systems. The current study therefore includes three languages as L1 background (Hungarian, Japanese and Dutch), which are distinctive in terms of orthographic systems and typology.

### **2.3. SYNTACTIC CHARACTERISTICS OF HUNGARIAN, JAPANESE, DUTCH AND THE ARTIFICIAL LANGUAGE IN LLAMA\_F (GRAMMATICAL INFERENCING)**

Hungarian is member of the Finno-Ugric language family, whilst most of European languages are categorized as Indo-European languages (Kiss, 2002). One of the distinctive aspects of the Finno-Ugric languages is its agglutinative characteristics. In agglutinative languages,

---

1 The non-Latin alphabetic group includes the L1 speakers of Chinese ( $n = 13$ ), Arabic ( $n = 2$ ), Greek ( $n = 1$ ) and Korean ( $n = 1$ ), the group with the Latin alphabetic group contains: French ( $n = 3$ ), Welsh ( $n = 2$ ), Spanish ( $n = 3$ ), Norwegian ( $n = 2$ ), German ( $n = 3$ ), Swiss-German ( $n = 1$ ), Danish ( $n = 1$ ), Yoruba ( $n = 1$ ) and Italian ( $n = 2$ ).

words contain multiple morphemes that mostly remain unchanged. It is also noteworthy that Hungarian has Subject-Verb-Object (SVO) as its main structure, followed by SOV (Kas et al., 2016). Hungarian, however, is often said to be a free word order language due to the variety of structures: SVO, SOV, OVS, OSV, VSO and VOS (Kiss, 2002).

Like Hungarian, Japanese is categorized as an agglutinative language. It is said to be part of the Altaic language family, but the exact origin of Japanese has not been determined (Maeda, 2007). The basic structure of Japanese is SOV, but the word order is usually flexible. With respect to verb position, the verb is placed at the end of a sentence. Japanese verbs have no personal conjugation, but they conjugate accompanying information such as tense. The Japanese sentence (1) is an example of a sentence without a subject and with the conjugation of the verb, showing the agglutinative nature of the language:

- (1) 答えさせられたくなかったら  
Kotae- sase- rare- tak- na- katta- ra  
Answer CAUS PASS DES NEG PST COND<sup>2</sup>  
'If (you) don't want to be made to answer'  
(Hasegawa, 2018, p. 4)

Dutch belongs to the West-Germanic group of Indo-European languages. Dutch is closely related to English and German and has two syntactic structures, SVO and SOV (Bennis & Israel, 2010). SVO mostly appears in the main clause, whilst SOV is found in subordinate clauses. Instead of grammatical cases, relations between words are indicated by using pronouns mostly (van der Sijs, 2005), and unlike Hungarian and Japanese, Dutch is not agglutinative in nature.

Upon analysing the sentences in the LLAMA\_F (grammatical inferencing) test, it was found that the syntax of this artificial language has a rich morphology, using inflection with adpositions. For instance, the sentence '*inut-ek ipot-arap*' is paired with a picture of 'two red rounded objects on a board'. The sentence would be broken down as follows:

*Inut-ek ipot-arap*

*inut*: on a board or above something

*-ek*: two objects

*ipot*: red colour

*arap*: rounded form

A morphological element '-ek' follows the adposition 'inut', which can be interpreted that the adposition has a characteristic of inflection. This suggests that the artificial language might be similar to an agglutinative language such as Hungarian in which postpositions can be attached to the stem of the word.

#### 2.4. ORTHOGRAPHIC CHARACTERISTICS OF HUNGARIAN, JAPANESE, DUTCH, AND THE SYSTEMS USED IN LLAMA\_B (VOCABULARY LEARNING)

Hungarian and Dutch are written using the Latin alphabet, which is a phonogram writing system. The two Japanese syllabic alphabets (*hiragana* and *katakana*) are also categorized as phonogram systems, whilst Japanese *kanji* is regarded as a logogram because each character of *kanji* depicts a specific meaning (Coderre et al., 2008; Tanaka, 2015). The most distinctive aspects of Japanese compared to both other languages in the present study may therefore be the number of writing systems used: Japanese uses two syllabic phonogram systems and one logographic writing system.

The Japanese writing system motivated the reinvestigation of the language neutrality of the LLAMA test because of its combination of orthographies. It is noteworthy that using both *kana* and *kanji* could affect cognitive processing (Coderre et al., 2008; Sakurai et al., 2000) because syllabic writing systems connect visual forms (graphemes) to sounds (phonemes), whilst logographic writing systems join forms (characters) with meanings (morphemes) and sounds (phonemes) (Coderre et al., 2008). It has also been shown that phonograms and logograms are

---

2 causative (CAUS), passive (PASS), desiderative (DES), negative (NEG), past (PST), conditional (COND).

processed differently (Tanaka, 2015). Such differential processing may transfer when reading in the L2, such that L2 learners with an alphabetic L1 background and L2 learners with non-alphabetic L1 backgrounds rely on different types of information during lexical processing in the L2 (Wang et al., 2003). Moreover, L1 speakers with logographic writing systems tend to rely more on visual information than on phonological information, while on the other hand, L1 readers with alphabetic language backgrounds tend to directly analyse phonological information (Wang & Geva, 2003). Because the writing experience in the L1 affects decoding and learning a new vocabulary in the L2 (Hamada & Koda, 2008), a language learning aptitude test may (unwantedly) be affected by the writing system used. L2 learners whose L1 writing system is closely related to that of the L2 are able to perform better than those who have less similar L1 language backgrounds (de Groot et al., 2002).

In previous validation studies, Chinese was included as a non-alphabetic language (Granena, 2013; Rogers et al., 2016, 2017). Although the ascendant of Japanese *kanji* is the Chinese character *hanji*, it is important here to emphasize that they have developed differently and they should not be classified in the same category (Hashimoto et al., 2017). The Japanese *kanji* is more graphic than the recent simplified Chinese characters mainly used in the mainland, and the written forms do not always match. Therefore, these writing systems should not be regarded as one type of logogram (Yokoyama, 2016). Furthermore, the Chinese *hanji* has only one mora and one pronunciation per character, by contrast, each Japanese *kanji* has one to three morae and it usually has more than two pronunciations. These differences argue for a separate investigation of Japanese as an L1. In addition to this need to study Japanese separately from Chinese, the L1 Japanese speakers may perform the LLAMA\_B (vocabulary learning) subtest in a more efficient way because they are accustomed to processing two different writing systems simultaneously. In the LLAMA\_B subtest, a picture (graphic) and a word (spelled in the Latin alphabet, thus phonographic) have to be matched and memorized, thus a graphic and phonographic system need to be processed at the same time. Given that the L1 Japanese speakers are used to processing two different writing systems in which one is graphic and the other phonographic, they may be found to score higher on this subtest.

### 3. RESEARCH QUESTIONS AND HYPOTHESES

The present study investigates the validity with respect to language neutrality of the LLAMA test, taking into account differences in syntactic typology and in writing systems. It focuses on two LLAMA subtests and uses one LLAMA subtest as a control test. To maximize potential contrasts between the agglutinative and non-agglutinative groups, whilst keeping the groups homogeneous and maximally comparable, the following L1 participant groups were chosen:

Agglutinative language groups:

- L1 Hungarian speakers who are majoring in Japanese
- L1 Japanese speakers who are majoring in Hungarian

Non-agglutinative language group

- L1 Dutch speakers who have never studied any agglutinative language majoring in a Romance language

Regarding the writing systems, Hungarian and Dutch both use the Latin alphabet (a so-called phonogram writing system), whereas Japanese uses two syllabic alphabets and a logographic writing system.

The abovementioned differences between Hungarian, Japanese and Dutch in terms of syntax and writing systems led to the following research questions and hypotheses:

- (a)** To what extent do speakers with L1 agglutinative language backgrounds outperform speakers with non-agglutinative L1 backgrounds on the LLAMA\_F (grammatical inferencing) test?

Since the LLAMA\_F subtest was found to have agglutinative characteristics, we hypothesize that the L1 Japanese and L1 Hungarian speakers will outperform the L1 Dutch speakers.

**(b)** To what extent do speakers with an L1 background that consists of multiple writing systems (phonogram and logogram) outperform speakers with L1 backgrounds using a single writing system (only phonogram) on the LLAMA\_B (vocabulary learning) test?

Given that the L1 Japanese speakers are used to processing two different writing systems, we hypothesize that they will outperform L1 Dutch participants and likely the L1 Hungarian speakers (who are learning Japanese). In addition, because the L1 Hungarian participants have started to learn a logographic writing system, they may outperform the L1 Dutch group.

Finally, the LLAMA\_D subtest was added as a control. Because research has indicated that the LLAMA\_D (sound recognition ability) is a measure for implicit learning ability (Granena, 2013), we expect to find no differences between the three language groups for this subtest.

Because earlier research (Rogers et al., 2017) has shown that prior L2 instruction is a significant variable in especially the LLAMA\_B and LLAMA\_F subtests, we added the number of languages known as a covariate in this study, serving as a proxy for amount of instruction in second (foreign) languages.

## 4. METHOD

### 4.1. PARTICIPANTS

Three groups of university students were recruited. All participants received a small gift as a token of appreciation for their participation. The first group consisted of Hungarian students (studying at Eötvös Loránd University or Károli Gáspár University) majoring in Japan Studies. The second group consisted of Japanese students (studying at Osaka University) majoring in Hungarian language and culture. The third group consisted of Dutch students (studying at Leiden University) majoring in a Romance language (e.g., French language and culture or Latin American Studies). Recruitment was initiated by e-mails to the relevant language teachers at the four universities. These teachers shared the call for participation in class or on online platforms. Additionally, the first author recruited participants in situ, by asking students just before or after class to participate. In total, 86 participants were recruited. Due to incomplete datasets or because of conditions that were not met, however, seven participants were excluded from the analyses. For instance, one of the Dutch students reported having learned Finnish, an agglutinative language, so this participant's data of this student were not included. **Table 1** shows the background information of all included participants in the three groups. All participants were informed of the purpose of the study and signed a consent form. Data were anonymized prior to data storage and analysis.

BACKGROUND INFORMATION	L1 HUNGARIAN (n = 25)	L1 JAPANESE (n = 30)	L1 DUTCH (n = 24)
Major	Japanese studies	Hungarian	Romance language (e.g., Latin American Studies, Italian Language and Culture)
Mean age (SD and range)	21.6 (SD = 2.10, range = 19–28)	19.6 (SD = 1.26, range = 18–23)	21.0 (SD = 2.96, range = 17–28)
Gender	7 M/18 F	4 M/26 F	3 M/21 F
Mean number of FL's (SD and range)	3.1 (SD = 0.93, range = 2–5)	2.4 (SD = 0.72, range = 2–5)	3.9 (SD = 1.28, range = 2–7)
Foreign languages reported	English, German, Japanese, Korean, Chinese, Finnish, Swedish	English, Hungarian, German, Korean, Chinese, French, Spanish, Italian, Polish, Pali, Bantu	English, German, French, Spanish, Italian, Portuguese, Maya, Russian, Lithuanian, Greek, Chinese

**Table 1** Background information of the participants in the three groups.

### 4.2. MATERIALS

#### 4.2.1. Language aptitude tests

The .exe-versions of the LLAMA\_B, \_D, and \_F tests were downloaded from <http://www.lognostics.co.uk/tools/llama/> onto a desktop or laptop and executed.

The LLAMA\_B was designed to measure the capability to learn and memorize new vocabulary items within a relatively short time. The LLAMA\_B consists of a study phase and a test phase. In the study phase, participants have two minutes to study a combination of 20 abstract pictures and corresponding words on the screen. These words are adopted from existing Central American languages (Meara, 2005) and presented with the Latin alphabet. Participants are required to learn the combinations of pictures and corresponding words within the given time frame. In the test phase, participants see one of 20 words on the screen and they are asked to choose the corresponding picture by clicking on the possible picture. Participants receive five points for each correct answer; the total scores of the LLAMA\_B can range between 0 to 100 (thus, the score is calculated as a percentage).

The LLAMA\_D test was developed to measure the ability to recognize repeated sounds in spoken languages, which is an essential ability for learning a foreign language (Service, 1992; Skehan, 2002; Speciale et al., 2004). The LLAMA\_D is designed to make it as language neutral as possible. The produced words are based on names, flowers and other natural objects in a British Columbian Indian language and these sounds were synthesized by using text transformer (AT&T Natural Voices) to make it difficult to recognize the sounds (Meara, 2005). In this test, a set of ten words in an unfamiliar language is played only one time. After that, participants need to click the arrow in the middle of the screen to perform the test. There is no study phase in this test. Compared to other LLAMA subtests, the LLAMA\_D barely invokes analytical abilities and could be considered as an implicit learning task (Granena, 2013). In the test phase, participants must indicate whether they have just heard the sound in the string of sounds by clicking on an icon on the screen. The sound recognition test consists of 30 questions and the obtained score is indicated on the screen. The score is converted by subtracting points for incorrect answers, and the highest possible score of the LLAMA\_D test is 75.

The LLAMA\_F was developed to measure explicit inductive learning ability, in other words, the analytical ability to extract the grammar rules of an unknown language from given information or materials. In this test, participants must discover the grammar rules of an artificial language in the study phase lasting five minutes. They are allowed to take notes. There are 20 buttons on the screen and by clicking one of them a combination of one sentence and one picture is displayed. This artificial language has a number of characteristics that might be difficult for L1 English speakers and it is based on languages such as Welsh (e.g., postpositions ‘-sa’ to show ‘one person’) (according to Meara, personal communication). In the test phase, two sentences and one picture are displayed at the same time and participants must choose one corresponding sentence. There is no time limit in the test phase. For correct answers five points are added, and as in the LLAMA\_D subtest, five points are subtracted from the score for each wrong answer. The total score ranges from 0 to 100.

#### 4.2.2. Questionnaire

Participants were asked to give information about their gender, age and the languages that they have learned (including a self-reported evaluation on their proficiency in these languages).

#### 4.2.3. Informal interview

After participation, a subset of participants was asked to reflect on their performance and potential strategies that they used. We aimed to ask mainly participants with relatively high scores and determined ‘relatively high’ according to the score interpretation indicated in the manual of the LLAMA test (Meara, 2005). For LLAMA\_B and LLAMA\_F, the manual says the scores between 75–100 are interpreted as “outstandingly good scores” (Meara, 2005, pp. 6, 17). Based on this interpretation, we have interviewed a selection of participants who were assessed as performing ‘outstandingly good’ on either LLAMA\_B or LLAMA\_F. Because of time constraints and practical constraints, for the Hungarian participants the selection was not based on the scores. Since the interview was held in an informal way, no set list of questions was used. However, the questions always concerned the following themes: (1) What are your thoughts on the subtest? (2) How did you work on the subtest? Did you use a strategy? And with respect to LLAMA\_F (grammatical inferencing), we also asked a question along the following lines: (3) Did you notice characteristics of the artificial language in the subtest? Specifically, do you think the artificial language is similar to any existing languages?



### 4.3. PROCEDURE

Each participant first read and signed the consent form and filled out the language background questionnaire. The first author then provided instructions about the experiment (orally), using a script based on the LLAMA manual (Meara, 2005). For carrying out the aptitude tasks, we used the same order as Saito (2017): LLAMA\_D, then \_B, and finally \_F. This order ensures that the most implicit task (LLAMA\_D, sound recognition) is carried out first, such that potential conscious strategies are less likely to be used. Most participants carried out the experiment in a quiet environment, such as a classroom or computer room. However, this was not possible for a couple of L1 Hungarian and L1 Dutch participants; they carried out the tasks in a common room. However, all participants used headphones during the experiment for LLAMA\_D and LLAMA\_B (vocabulary learning). To be able to make notes, participants received pen and paper before the start of LLAMA\_F (grammatical inferencing). Scores for all the tests were recorded by the software and noted down by the first author. For 16 students, the first author asked them in an informal interview how they had proceeded in carrying out the task. The answers to these questions were written down by the first author. The entire experiment including the interview after the test lasted around 30 minutes.

### 4.4. DATA ANALYSIS

To test whether there were differences between the three groups, we carried out a MANCOVA, with the three scores for the LLAMA tests as dependent variables, language group as the independent factor (with three levels) and number of languages known as covariate. When checking the assumptions for running a MANCOVA, because some participants noted knowing eight languages but none noted seven, we rank-transformed the number of languages known. Three Levene's tests showed that for the dependent variables, homogeneity of variance could be assumed ( $p$ 's > 0.19), and Box's M test revealed that equality of covariance could be assumed ( $p = 0.91$ ). Finally, no anomalies were found when analysing the standardized residuals. Whenever a significant difference was indicated for one of the dependent variables, we carried out post-hoc t-tests to check which groups differed from each other. Taking into account the group sizes, the analyses would reach a power of 0.88, if we assumed the effects to be large ( $f = 0.4$ ). However, if we assume effects to be medium or small (thus given an  $f$  of 0.25 or 0.1), power in this study is calculated to reach only 0.48 or 0.11. The quantitative data are available on this webpage: <https://osf.io/ve7xa/>. The notes made during the informal interviews were summarized, and exemplary findings from these summaries are reported in the next section.

## 5. RESULTS

### 5.1. EXPLORING DIFFERENCES IN APTITUDE SCORES

**Table 2** shows the means and standard deviations of the original scores on the LLAMA tests by the three participant groups. Boxplots of these data can be found in **Figure 1**. As explained, the covariate number of languages known was rank-transformed prior to analyses.<sup>3</sup> **Table 3** shows the estimated marginal means and standard errors of all subtests and groups after correcting for (rank) number of languages known.

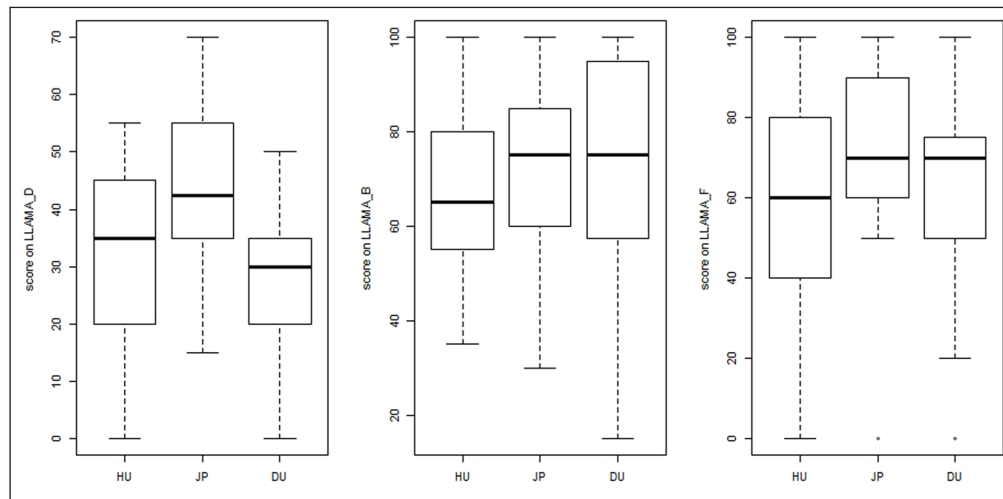
L1	LLAMA_D	LLAMA_B	LLAMA_F
Hungarian ( $n = 25$ )	33.40 (15.99)	66.00 (18.48)	59.80 (24.26)
Japanese ( $n = 30$ )	43.83 (12.78)	72.17 (19.77)	71.50 (20.14)
Dutch ( $n = 24$ )	26.88 (13.58)	72.29 (23.91)	62.08 (23.59)

**Table 2** Means (SD) of scores on the three aptitude tests.

The MANCOVA showed an overall effect of language group ( $F(6, 148) = 3.38, p = 0.004$ ). From the separate analyses, it was found that there were group differences for LLAMA\_D, sound recognition ability ( $F(2, 75) = 8.77, p < 0.001, \eta_p^2 = .19$ ) and LLAMA\_F ( $F(2, 75) = 3.89, p = 0.025, \eta_p^2 = 0.09$ ), but not for LLAMA\_B, vocabulary learning ( $F = 1.79, p = 0.174$ ). For the

<sup>3</sup> When carrying out the analysis with the original scores for number of languages known as a covariate, the same results were found.

LLAMA\_B analysis, the number of languages known turned out to be a significant predictor ( $F(1, 75) = 4.97, p = 0.029, \eta_p^2 = 0.06$ ) and in the LLAMA\_F (grammatical inferencing) analysis, the number of languages as covariate approached significance ( $F(1, 75) = 3.74, p = 0.057, \eta_p^2 = 0.05$ ). Pairwise comparisons revealed that for LLAMA\_D, Japanese outperformed both Hungarian ( $p = 0.005$ ) and Dutch ( $p < 0.001$ ) participants, but Hungarian and Dutch participants did not show differential scores ( $p = 0.074$ ). For LLAMA\_F, pairwise comparisons revealed that Japanese outperformed both Hungarian ( $p = 0.014$ ) and Dutch ( $p = 0.020$ ), but Hungarian and Dutch were not different from each other ( $p = 0.85$ ). The total adjusted explained variance ( $R^2$ ) for the models for LLAMA\_D and LLAMA\_F were 0.19 and 0.06, respectively.



**Figure 1** Boxplots of scores on the LLAMA\_D, LLAMA\_D, and LLAMA\_F subtests by the three L1 background groups (Boxes include 25%–75% of the data. Horizontal lines mark medians. Whiskers are  $\pm 1.5^*$  interquartile range. The dots are outliers below or above the interquartile range).

L1	LLAMA_D		LLAMA_B		LLAMA_F	
	EST. MARG. MEANS (SE)	CI	EST. MARG. MEANS (SE)	CI	EST. MARG. MEANS (SE)	CI
Hungarian	33.19 (2.83)	[27.54, 38.83]	65.26 (4.06)	[57.18, 73.35]	59.10 (4.45)	[50.24, 67.96]
Japanese	45.03 (2.89)	[39.28, 50.78]	76.32 (4.13)	[68.08, 84.55]	75.45 (4.53)	[66.42, 84.48]
Dutch	25.60 (3.20)	[19.23, 31.97]	67.87 (4.58)	[58.75, 76.99]	57.87 (5.02)	[47.88, 67.87]

**Table 3** Estimated Marginal Means (SE) of scores on the three aptitude tests, including Confidence Intervals (CIs).

## 5.2. EXPLORING DIFFERENCES IN STRATEGIES

Informal interviews were held with a selection of participants, most of whom had scored relatively high on either LLAMA\_B or LLAMA\_F. In total, four L1 Hungarian, eight L1 Japanese, and four L1 Dutch participants answered informal questions immediately after finishing LLAMA\_F.<sup>4</sup>

All 16 participants reported that LLAMA\_D (sound recognition) had been particularly demanding but did not mention any conscious strategies they used. Regarding LLAMA\_B (vocabulary learning), the most used (self-reported) strategy was to mentally connect the “words” they read to similar sounding words from their L1 (or an L2). For instance, L1 Hungarian speakers connected “CIB” to the existing commercial bank CIB. Some L1 Hungarian speakers who were learning Japanese as an L2 specifically mentioned that they tried to use similar strategies they had been using for learning Japanese *kanji*. L1 Japanese speakers memorized “KABAN” (*bag* in Japanese) with the accompanying picture of a log by remembering “a wooden bag”. Similarly, the L1 Dutch speakers noted that they used their L2 English to associate “MEN” with the picture of a little man. Furthermore, an L1 Dutch speaker mentioned resemblance of the words used in the LLAMA\_B to a Mayan language. The participant thus made use of personal knowledge of a Mayan language to complete this subtest.<sup>5</sup>

<sup>4</sup> As mentioned in the Method section, because of time constraints and practical constraints, we were not able to only select high-scoring participants for the Hungarian group. One of the Hungarian participants did have a ‘relatively high’ score (90) on both LLAMA\_B and LLAMA\_F. The other three scored at least 60 on either the LLAMA\_B or LLAMA\_F subtest.

<sup>5</sup> Excluding this participant from the inferential statistical analyses did not lead to any changes in conclusions drawn from these analyses.

Finally, regarding LLAMA\_F (grammatical inferencing), the L1 Japanese speakers noted that the artificial language resembled Hungarian, especially in the way that the artificial language used affixes. L1 Hungarian speakers noted that the artificial language resembled an agglutinative language. In contrast, the L1 Dutch speakers mentioned no resemblances of the artificial language to a language that they knew.

## 6. DISCUSSION AND CONCLUSION

Several studies have investigated the extent to which the LLAMA language aptitude tests can be considered language neutral (Granena, 2013; Rogers et al., 2016, 2017). So far, comparing a number of L1 backgrounds, including those using different writing systems such as Arabic and Mandarin, no significant differences between participants with different language backgrounds have been found. However, until now, neither agglutinative language backgrounds nor L1 backgrounds in which multiple writing systems are used have been taken into account. Therefore, this study selected participants from three different L1 backgrounds: Dutch (non-agglutinative, phonogram/Latin alphabet), Hungarian (agglutinative, phonogram/Latin alphabet), and Japanese (agglutinative, phonogram/syllabic alphabet and logogram/Japanese kanji). The three groups were otherwise comparable: All participants were university students majoring in language and culture. All Japanese students studied Hungarian, all Hungarian students studied Japanese, and all Dutch students studied a Romance language (and had never learnt any agglutinative language).

Three subtests of the LLAMA aptitude battery were administered: LLAMA\_D (an implicit learning test gauging the ability to recognize new sounds), LLAMA\_B (a vocabulary learning test that uses icons combined with a Latin script) and LLAMA\_F (a grammar inferencing test that includes agglutinative characteristics). Due to the agglutinative characteristics in LLAMA\_F, the combination of icons and Latin script in LLAMA\_B and the implicitness of LLAMA\_D, we hypothesized (a) that Japanese and Hungarian participants would outperform the Dutch participants in LLAMA\_F, (b) that Japanese speakers would outperform Dutch speakers and would possibly outperform Hungarian in LLAMA\_B and (c) that no differences would be found for LLAMA\_D.

For the first hypothesis concerning LLAMA\_F (grammatical inferencing), results showed that the Japanese participants outperformed both groups. Although we expected the Japanese participants to outperform Dutch participants, because they have experience with two agglutinative languages, we cannot explain why the Japanese participants outperformed the Hungarian participants. The L1 Hungarian speakers likewise have an agglutinative language as L1 background and are majoring in language and culture of another agglutinative language — Japanese. The informal interview did reveal that both Japanese and Hungarian participants noted that the grammar was similar to their own language or to the language they were learning.

Regarding the second hypothesis, results showed that there were no differences between the participant groups with respect to the scores on LLAMA\_B (vocabulary learning). This finding confirms the conclusions from the studies by Granena (2013) and Rogers et al. (2016, 2017) that LLAMA\_B can be seen as a language-neutral aptitude test. The conclusion can now be broadened to include a language such as Japanese, which uses multiple writing systems.

Finally, the hypothesis regarding LLAMA\_D (sound recognition) was also not borne out. Contrary to our expectations that this part of the LLAMA-test would be language neutral, the Japanese outperformed both Dutch and Hungarian. No further differences between the groups were found. This test has been argued to be a test of implicit learning and therefore we hypothesized that no differences between the three groups would occur. However, considering the task of recognizing novel arbitrary sounds, it may be the case that L1 Japanese speakers were more apt at listening for and recognizing new sounds because Japanese is known as a language with a very rich inventory of sound-symbolic or mimetic words (Hamano, 1998; Kita, 1997). For instance, Japanese uses not only onomatopoeia, but also mimetic words for emotions or states, such as being excited ('wakuwaku'), in love ('kyunkyun'), fidgety ('sowasowa') or edgy ('iraira'). They are expressed by sounds that are repeated and their sound-to-meaning correspondence is not (perceived as) arbitrary. Hirata (2013) says that Japanese speakers will typically attach an impression to a sound. Because of the large inventory of mimetic words and onomatopoeic

words that L1 Japanese speakers have, a new sound is likely to induce a specific impression or semantics. From this, we may speculate that participants with Japanese as an L1 background are more attuned to recognizing new sounds, as they have possibly already related the sound to (a kind of) meaning when hearing it the first time. There is no confirmation of this hypothesis in the comments made by the participants in the informal interview. However, such a lack of comments is to be expected if the LLAMA\_D does indeed mainly tap into implicit learning. An alternative explanation for our results may be that the L1 Japanese group scored higher due to an unknown general superiority in language aptitude abilities or another covariate not included in our questionnaire, leading them to have a higher aptitude in general. It is worth noting, however, that no differences were found with respect to the LLAMA\_B subtest, which indicates that this unknown general confound would therefore not hold for all subtests, or at least not to the same degree.

To account for possible differences in the L1 groups with respect to amount of L2 instruction received, we added the number of languages known, as reported by our participants, as a covariate. This covariate was significant in the analyses of the LLAMA\_B (vocabulary learning) and approached significance in the analyses of LLAMA\_F (grammatical inferencing), which confirms conclusions from earlier research (Rogers et al., 2017), who explicitly investigated whether receiving L2 instruction would lead to advantages on the LLAMA test.

As mentioned in the Method section, the power calculation showed that the current study was suitably sensitive to large differences between groups (L1 backgrounds only). This may be specifically troublesome in the case where research is attempting to ‘prove’ a null hypothesis, as is the case in the current study and in similar studies comparing participant scores with different language backgrounds. Such studies would conclude a test to be valid if there were no significant differences between the groups. Another potential limitation of the study might be the fact that gender was not balanced in our study: In all three groups, female participants outnumbered male participants. This means that the results of this study can only be generalized to a similar population.


We conclude that the results of the current study have implications for the use of the LLAMA tests, especially when participant scores are compared across different language backgrounds. First of all, the amount of instruction and the number of languages that the participants know need to be taken into account when comparing participants. The more instruction or languages known, the higher participants will score on LLAMA\_B (this study and Rogers et al., 2017) and on LLAMA\_F (Rogers et al, 2017; marginal effect in this study). Secondly, we found no differences between the different L1 backgrounds regarding the scores on the LLAMA\_B (vocabulary learning) subtest, which suggests that LLAMA\_B is indeed language neutral. From the differences we found on the scores of the LLAMA\_D (sound recognition) and LLAMA\_F (grammatical inferencing) subtests between the groups, no firm conclusions can be drawn yet. With respect to the LLAMA\_D subtest, we speculated that for participants who have Japanese as an L1, a language that uses many non-arbitrary sound-to-meaning mappings, recognizing new sounds in a new language may be easier. This speculation needs to be investigated in future studies. If the finding were to hold, it would mean that participants who were more attuned to listening to new sounds because of their L1 background would indeed perform better in the LLAMA\_D subtest. Likewise, the differences we found for the LLAMA\_F subtest also warrant further research. If it is the case that knowing agglutinative languages gives learners an advantage for learning another language with agglutinative characteristics, such as the artificial language used in LLAMA\_F, one could argue that the LLAMA\_F subtest is not entirely language neutral. However, since we found that the L1 Japanese participants outperformed not only L1 Dutch participants (non-agglutinative) but also the L1 Hungarian speakers (agglutinative language), no firm conclusion can be drawn. We can only conclude that L1 Japanese speakers may be at an advantage for both the LLAMA\_F and LLAMA\_D subtests.<sup>6</sup>


---

<sup>6</sup> In April 2020, a new version of the LLAMA test was published (Meara & Rogers, 2020). The fact that this version is even more easily accessible, and still free to use, will no doubt inspire new research using the LLAMA test, including research investigating the language neutrality of this new version. As one of the reviewers pointed out, a number of issues that have been raised with the version of the LLAMA tests used in the current study have been corrected in this latest version of the online LLAMA tests (e.g., an issue with incorrect coding in some items of the LLAMA\_F subtest, see also Bokander & Bylund, 2020); how LLAMA\_D and LLAMA\_F are scored, correcting for wrong answers yet being a binary choice compared to LLAMA\_B, which has a much lower guessing possibility).

The research was not sent out for approval by an ethics committee, which was in line with local regulations for student research at time of data collection. Prerequisites as asked by the current local ethical committee (faculty of humanities, Leiden University) were taken into account: Participants gave their informed consent, and all data were analysed, stored, and reported on anonymously.

## AUTHOR AFFILIATIONS

**Momo Mikawa**  [orcid.org/0000-0002-2237-6143](https://orcid.org/0000-0002-2237-6143)  
Leiden University Center for Linguistics, Leiden University, NL

**Nivja H. De Jong**  [orcid.org/0000-0002-3680-3820](https://orcid.org/0000-0002-3680-3820)  
Leiden University Center for Linguistics, Leiden University, NL;  
ICLON Graduate School of Teaching, Leiden University, NL

## REFERENCES

- 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. DOI: <https://doi.org/10.1016/j.lindif.2016.06.023>
- Bennis, H., & Israel, T.** (2010). *Syntaxis van het Nederlands* [Syntax of the Dutch language]. Amsterdam University Press.
- Bokander, L., & Bylund, E.** (2020). Probing the internal validity of the LLAMA language aptitude tests. *Language Learning, 70*, 11–47. DOI: <https://doi.org/10.1111/lang.12368>
- Carroll, J. B.** (1973). Implications of aptitude test research and psycholinguistic theory for foreign-language teaching. *Linguistics, 11*(112), 5–14. DOI: <https://doi.org/10.1515/ling.1973.11.112.5>
- Carroll, J. B.** (1981). Twenty-five years of research on foreign language aptitude. In K. C. Diller (Ed.), *Individual differences and universals in language learning aptitude* (pp. 83–118). Newbury House.
- Carroll, J. B., & Sapon, S. M.** (1959). *Modern Language Aptitude Test*. Psychological Corporation.
- Coderre, E. L., Filippi, C. G., Newhouse, P. A., & Dumas, J. A.** (2008). The Stroop effect in kana and kanji scripts in native Japanese speakers: An fMRI study. *Brain and language, 107*(2), 124–132. DOI: <https://doi.org/10.1016/j.bandl.2008.01.011>
- de Groot, A. M., Borgwaldt, S., Bos, M., & van den Eijnden, E.** (2002). Lexical decision and word naming in bilinguals: Language effects and task effects. *Journal of Memory and Language, 47*(1), 91–124. DOI: <https://doi.org/10.1006/jmla.2001.2840>
- Dörnyei, Z.** (2014). *The psychology of the language learner: Individual differences in second language acquisition*. Erlbaum. DOI: <https://doi.org/10.4324/9781410613349>
- 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* (pp. 589–630). Blackwell. DOI: <https://doi.org/10.1002/9780470756492.ch18>
- Granena, G.** (2013). 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. DOI: <https://doi.org/10.1075/llt.35.04gra>
- Granena, G., & Long, M. H.** (2013). Age of onset, length of residence, language aptitude, and ultimate L2 attainment in three linguistic domains. *Second Language Research, 29*(3), 311–343. DOI: <https://doi.org/10.1177/0267658312461497>
- 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. DOI: <https://doi.org/10.1111/0026-7902.00076>
- Hamada, M., & Koda, K.** (2008). Influence of first language orthographic experience on second language decoding and word learning. *Language Learning, 58*(1), 1–31. DOI: <https://doi.org/10.1111/j.1467-9922.2007.00433.x>
- Hamano, S.** (1998). *The sound symbolic system of Japanese*. CSLI Publications & Kuroshio.
- Hasegawa, Y.** (2018). Introduction. In Y. Hasegawa (Ed.), *The Cambridge Handbook of Japanese Linguistics* (pp. 3–14). Cambridge University Press. DOI: <https://doi.org/10.1017/9781316884461>
- Hashimoto, K., Uno, A., Sambai, A., & Mizumoto, G.** (2017). A patient with aphasia using the nonsemantic lexical route for Kanji reading. *Neurocase, 23*(5–6), 270–277. DOI: <https://doi.org/10.1080/13554794.2017.1390140>
- Hirata, S.** (2013). ‘Onseishouchou to kankakukanicchi no ruijisei’ [Similarities between sound symbolism and cross-modal correspondences]. In K. Shinohara & R. Uno (Eds.), *Onomatopoe-kenkyuu no syatei: Tikazuku oto to imi* [Sound symbolism and mimetics: Rethinking the relationship between sound and meaning in language] (pp. 101–115). Hituzi Syobo.

- Kas, B., Lukács, Á., & Szentkúti-Kiss, K.** (2016). A szórend és az esetjelölés szerepe specifikus nyelvi zavart mutató gyerekek mondatfeldolgozásában [The role of word order and case designation in sentence processing in children with specific language disorders]. In B. Kas (Ed.), *Szavad ne feledd! Tanulmányok Bánréti Zoltán tiszteletére* ["Remember your word!": Studies in honor of Zoltán Bánréti] (pp. 367–381). MTA Nyelvtudományi Intézet.
- Kepinska O., de Rover M., Caspers J., & Schiller N. O.** (2017). On neural correlates of individual differences in novel grammar learning: An fMRI study. *Neuropsychologia*, 98, 156–168. DOI: <https://doi.org/10.1016/j.neuropsychologia.2016.06.014>
- Kiss, K. É.** (2002). *The syntax of Hungarian*. Cambridge University Press. DOI: <https://doi.org/10.1017/CBO9780511755088>
- Kita, S.** (1997). Two-dimensional semantic analysis of Japanese mimetics. *Linguistics*, 35, 379–415. DOI: <https://doi.org/10.1515/ling.1997.35.2.379>
- Li, S.** (2016). The construct validity of language aptitude: A meta-analysis. *Studies in Second Language Acquisition*, 38(4), 801–842. DOI: <https://doi.org/10.1017/S027226311500042X>
- Linck, J. A., Hughes, M. M., Campbell, S. G., Silbert, N. H., Tare, M., Jackson, S. R., & Doughty, C. J.** (2013). Hi-lab: A new measure of aptitude for high-level language proficiency. *Language Learning*, 63(3), 530–566. DOI: <https://doi.org/10.1111/lang.12011>
- Maeda, T.** (2007). Classification of the Japanese language. In Y. Hida, Y. Endo, M. Kato, T. Sato, K. Hachiya & T. Maeda (Eds.), *The Research Encyclopedia of Japanese Linguistics* (pp. 24–37). Meijishoin Co.
- Meara, P.** (2005). *LLAMA language aptitude tests: The Manual*. Lognostics.
- Meara, P., Milton, J., & Lorenzo-Dus, N.** (2002). *Swansea language aptitude tests (LAT) v2.0*. Lognostics.
- Meara, P. M., & Rogers, V. E.** (2020). *The LLAMA Tests v3*. Lognostics.
- 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. DOI: <https://doi.org/10.1177/001316447603600216>
- Pimsleur, P.** (1966). *Pimsleur language aptitude battery (Forms)*. Harcourt, Brace and World Incorporated.
- Rogers, V. E., Meara, P., Aspinall, R., Fallon, L., Goss, T., Keey, E., & Thomas, R.** (2016). Testing aptitude. *EUROSLA Yearbook*, 16(1), 179–210. DOI: <https://doi.org/10.1075/eurosla.16.07rog>
- 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. DOI: <https://doi.org/10.22599/jesla.24>
- Saito, K.** (2017). Effects of sound, vocabulary, and grammar learning aptitude on adult second language speech attainment in foreign language classrooms. *Language Learning*, 67(3), 665–693. DOI: <https://doi.org/10.1111/lang.12244>
- Sakurai, Y., Momose, T., Iwata, M., Sudo, Y., Ohtomo, K., & Kanazawa, I.** (2000). Different cortical activity in reading of Kanji words, Kana words and Kana nonwords. *Cognitive Brain Research*, 9(1), 111–115. DOI: [https://doi.org/10.1016/S0926-6410\(99\)00052-X](https://doi.org/10.1016/S0926-6410(99)00052-X)
- Service, E.** (1992). Phonology, working memory, and foreign-language learning. *The Quarterly Journal of Experimental Psychology Section A*, 45(1), 21–50. DOI: <https://doi.org/10.1080/14640749208401314>
- Skehan, P.** (2002). Theorising and updating aptitude. In P. Robinson (Ed.), *Individual differences and instructed language learning* (pp. 69–93). Benjamins. DOI: <https://doi.org/10.1075/llt.2.06ske>
- Sparks, R. L., Ganschow, L., Fluharty, K., & Little, S.** (1995). An exploratory study on the effects of Latin on the native language skills and foreign language aptitude of students with and without learning disabilities. *The Classical Journal*, 91(2), 165–184.
- 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. DOI: <https://doi.org/10.1017/S0142716404001146>
- Tanaka, M.** (2015). Japanese Kanji word processing for Chinese Learners of Japanese: a study of homophonic and semantic primed lexical decision tasks. *Theory and Practice in Language Studies*, 5(5), 900–905. DOI: <https://doi.org/10.17507/tpls.0505.03>
- Tolchinsky, L., Levin, I., Aram, D., & McBride-Chang, C.** (2012). Building literacy in alphabetic, abjad and morphosyllabic systems. *Reading and Writing*, 25(7), 1573–1598. DOI: <https://doi.org/10.1007/s11145-011-9334-7>
- van der Sijs, N.** (2005). *De geschiedenis van het Nederlands in een notendop*. [The history of Dutch in a nutshell.] Prometheus.
- Wang, M., & Geva, E.** (2003). Spelling performance of Chinese children using English as a second language: Lexical and visual-orthographic processes. *Applied Psycholinguistics*, 24(1), 1–25. DOI: <https://doi.org/10.1017/S0142716403000018>
- Wang, M., Koda, K., & Perfetti, C. A.** (2003). Alphabetic and nonalphabetic L1 effects in English word identification: A comparison of Korean and Chinese English L2 learners. *Cognition*, 87(2), 129–149. DOI: [https://doi.org/10.1016/S0010-0277\(02\)00232-9](https://doi.org/10.1016/S0010-0277(02)00232-9)

- Yalçın, Ş., Çeçen, S., & Erçetin, G.** (2016). The relationship between aptitude and working memory: An instructed SLA context. *Language Awareness*, 25(1–2), 144–158. DOI: <https://doi.org/10.1080/09658416.2015.1122026>
- Yilmaz, Y., & Granena, G.** (2016). The role of cognitive aptitudes for explicit language learning in the relative effects of explicit and implicit feedback. *Bilingualism: Language and Cognition*, 19(1), 147–161. DOI: <https://doi.org/10.1017/S136672891400090X>
- Yokoyama, S.** (2016). 電子メディアの漢字と東アジアの文字生活 [Kanji in digital media and the ordinary use of characters in East Asia]. In Y. Liang (Ed.), *Propagation and Interchange of Transnational Culture in East Asia: Focusing on Media, Japanese Studies Research Series* (pp. 205–222). National Taiwan University Press. DOI: <https://doi.org/10.6327/NTUPRS-9789863501763>

**Mikawa and De Jong** 100  
*Journal of the European  
Second Language  
Association*  
DOI: 10.22599/jesla.71

TO CITE THIS ARTICLE:

Mikawa, M., & De Jong, N. H. (2021). Language neutrality of the LLAMA test explored: The case of agglutinative languages and multiple writing systems. *Journal of the European Second Language Association*, 5(1), 87–100. DOI: <https://doi.org/10.22599/jesla.71>

Submitted: 01 June 2020

Accepted: 29 June 2021

Published: 15 July 2021

COPYRIGHT:

© 2021 The Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (CC-BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited. See <http://creativecommons.org/licenses/by/4.0/>.

*Journal of the European Second Language Association*, is a peer-reviewed open access journal published by White Rose University Press.