Aptitude for Explicit and Implicit Learning

RENATO PAVLEKOVIC (ID)

KAREN ROEHR-BRACKIN (ID)

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

ABSTRACT

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

CORRESPONDING AUTHOR:

Renato Pavlekovic

University of Essex, United Kingdom

r.pavlekovic@essex.ac.uk

1. INTRODUCTION

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

2. BACKGROUND

2.1. LANGUAGE LEARNING APTITUDE AND WM AS PREDICTORS OF L2 ATTAINMENT

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

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

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

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

Empirical investigations into the relationship between WM and aptitude have led to mixed results. Several studies have identified no or weak relationships between measures of the

two constructs (Roehr & Gánem-Gutiérrez, 2009; Yoshimura, 2001); Li's (2016) meta-analysis identified a weak correlation between PWM and EWM and overall aptitude and aptitude components.

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

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

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

2.2. MEASURING APTITUDE

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

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

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

Taken together, these results could be interpreted as emerging evidence for a multi-componential structure of implicit aptitude (Li & DeKeyser, 2021), with LLAMA D potentially probing implicit memory ability in the verbal domain and the SRT task domain-general implicit learning ability (Granena, 2020). At the same time, seemingly inconsistent findings involving LLAMA D could be attributable to the test instructions used in any given study and thus be a methodological issue (Li, 2022). Specifically, if participants are informed that they will be tested on the items they hear in the exposure phase, attempts at intentional and therefore explicit learning could ensue. In order to test this hypothesis, Iizuka and DeKeyser (2023) compared three types of LLAMA D instructions ranging from more to less explicit ('listen and memorize', 'just listen', 'sound check')[1] and their effects on task performance. The researchers found that only the 'just listen' instructions that asked participants to carefully listen to the stimuli resulted in a relationship between LLAMA D and the SRT task. However, surprisingly, the relationship was negative. In an attempt to interpret this unexpected result, the researchers suggest that an ability to focus on the auditory stimuli helped with LLAMA D, but had the opposite effect in the case of the SRT task, where focusing on the stimuli on a trial-to-trial basis may have prevented successful (implicit) learning of the probabilistic sequence. Such an interpretation, in turn, might suggest that implicit aptitude is primarily a lack of interference rather than a measurable ability that enhances learning (Iizuka & DeKeyser, 2023). In this regard, the study reports a novel finding and offers an interesting interpretation that could potentially have wide-ranging implications for the conceptualization of implicit aptitude. However, replication is clearly needed.

3. THE CURRENT STUDY

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

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

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

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

---

1   As pointed out by a reviewer, none of these instructions correspond to the instructions used in Rogers et al. (2016) or the current LLAMA v.3.

3.1. METHOD

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

3.1.1. Participants

A total of 86 L1 Croatian learners of L2 English participated in the study. At the time of data collection, the participants had been learning English for between 6 and 13 years (*M* = 10, *SD* = 1.72) in the context of mandatory classes as a part of their school curriculum. The sample included 62 women, 22 men, and two participants who preferred not to disclose their gender. Participants were in secondary education and ranged in age from 15 to 18 years (*M* = 16.14, *SD* = 1.29).

3.1.2. Instruments and procedure

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

3.1.3. Explicit and implicit aptitude

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

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

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

---

2   The measures are available at https://osf.io/95uhk/?view_only=7df3be6c9bf44c768c5cee2efce2f544.

LLAMA F is a grammatical inferencing task in which participants have four minutes to work out the rules of an unknown language. During the learning phase, they click on buttons that reveal picture stimuli with corresponding written descriptions. There are 20 items, and participants can click on any button any number of times. During the untimed test phase which comprises 20 items, participants are presented with similar stimuli and must select a combination of words that correctly describes the picture at hand. Participants construct their answers from a board of 16 words. The version used in the present study was equivalent to v.3 except for the fact that all 20 items from v.1 were used. In our partial-credit scoring system, each correct word yielded up to two points: one point for the appropriate word itself, and one point if the word was in the correct position. The maximum score was 132.

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

3.1.4. WM

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

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

3.1.5. L2 proficiency

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

The Oxford Placement test (Oxford University Press, n.d.) was used to assess participants' L2 reading and listening comprehension. The reading section draws on test takers' knowledge of grammatical form and meaning, implied meaning and overall reading comprehension. The listening section assesses test takers' listening comprehension through ten dialogues of

varying lengths and five short monologues. Both sections are designed to test how well learners understand the meaning of what is being communicated as an indicator of general language ability (Oxford University Press, n.d.). The test is adaptive (i.e., the difficulty of presented items is kept in line with each learner's performance). The test provider's platform generates a total test score as well as separate scores for each section, with a maximum score of 120 for each section.

3.1.6. Data analysis

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

3.2. RESULTS

This section provides answers to our three research questions in chronological order.

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

To address this question, we first considered participants' overall performance on the LLAMA. Participants scored highest on LLAMA B (*M* = 57.18, *SD* = 22.34) and lowest on LLAMA D (*M* = 37.57, *SD* = 23.10). LLAMA E (α = 0.97) and LLAMA F (α = 0.95) showed excellent reliability; LLAMA B also showed very good reliability (α = 0.81). LLAMA D yielded a lower but still acceptable coefficient (α = 0.72). The full descriptive statistics are shown in Table A in the online supplementary materials.[3]

Additionally, we calculated participants' SRT task scores and scrutinized differences in mean RT between the eight task blocks to establish the time course of any learning effects. First, error responses (9% of the data) were discarded. Significant outliers (1% of the data), defined as values of more than three SDs from the mean RT for each participant in each block (Granena, 2016; Suzuki & DeKeyser, 2015), were likewise discarded, reducing the sample size to 83. Each participant's SRT score was calculated by subtracting the average RT in the training condition from the average RT in the control condition. Figure 1 shows mean RTs in each block of the SRT task on both training and control trials.
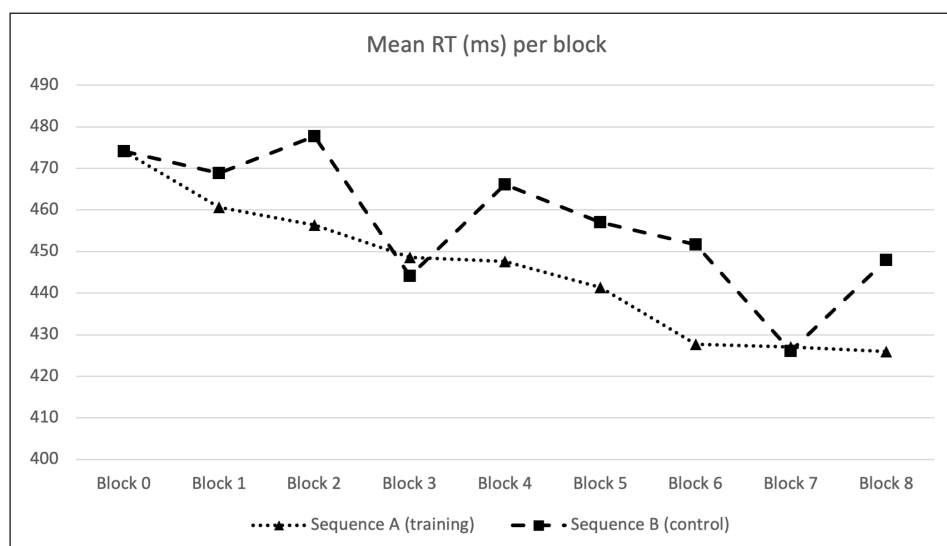
---

3   The supplementary materials are available at https://osf.io/385tf?view_only=4443aa5e41b34d22a6ca2075be0a9338.

As Figure 1 indicates, the RT in the control condition was larger than in the training condition on all blocks except for blocks 3 and 7. The mean RT for the training condition across all blocks was 447ms (SD = 67); the mean RT for the control condition was 457ms (SD = 71). Table B in the supplementary materials shows mean RTs broken down by block and condition.

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

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

Bivariate correlations (Spearman's rho) were run to examine whether there was any convergence between the SRT task and LLAMA D as hypothesized measures of implicit aptitude. Figure 2 shows correlation coefficients (upper triangle), scatterplots for variable pairs (lower triangle) and density plots for each variable (on the diagonal). The results show no significant association between SRT and LLAMA D, thus indicating divergence.[4]
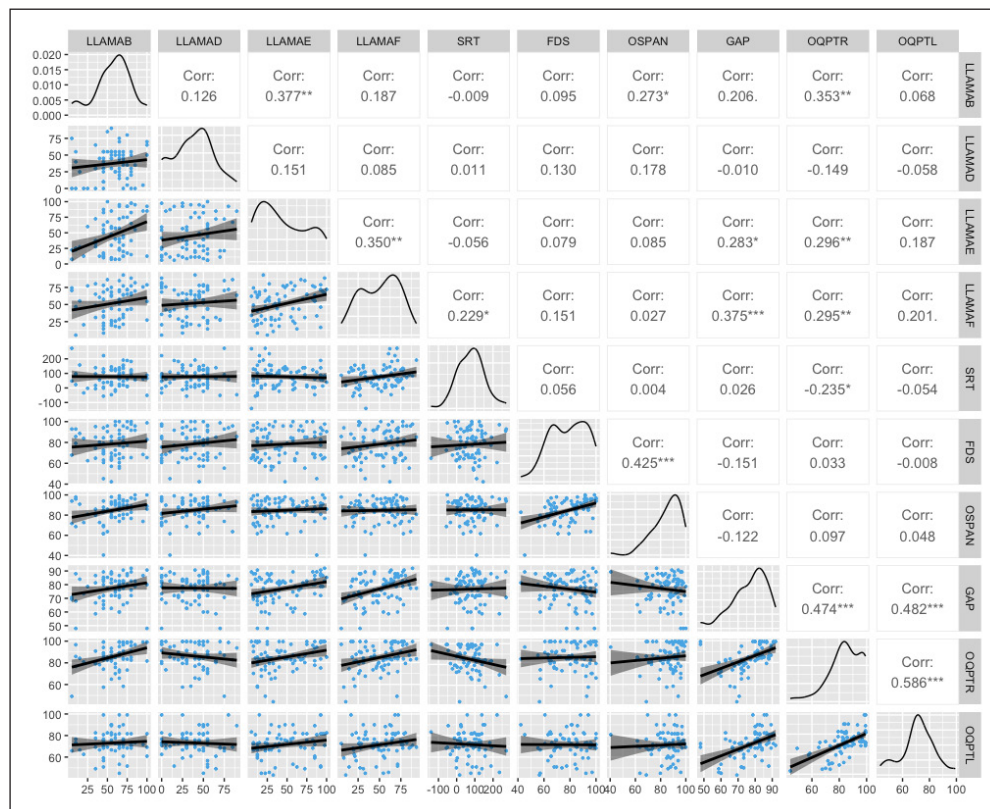

Figure 2 Correlations (Spearman's rho) between measures of aptitude, WM, and L2 proficiency.

Finally, we conducted an exploratory factor analysis, using principal component analysis with direct oblimin (oblique) rotation, following confirmation that underlying factors were related (Tabachnick & Fidell, 2014, p. 651). Assumptions were met, with a Kaiser-Meyer-Olkin

---

4   A reviewer suggested scoring only LLAMA D items which should attract a 'yes' answer. We also looked at RT scores for both old and new items. Neither of these scoring systems led to different results.

(KMO) measure of sampling adequacy of 0.55 and Bartlett's test of sphericity significant. The analysis yielded three components with eigenvalues above 1. The factor loadings are shown in Figure 3. A detailed overview of factor loadings can be found in Table D in the supplementary materials.
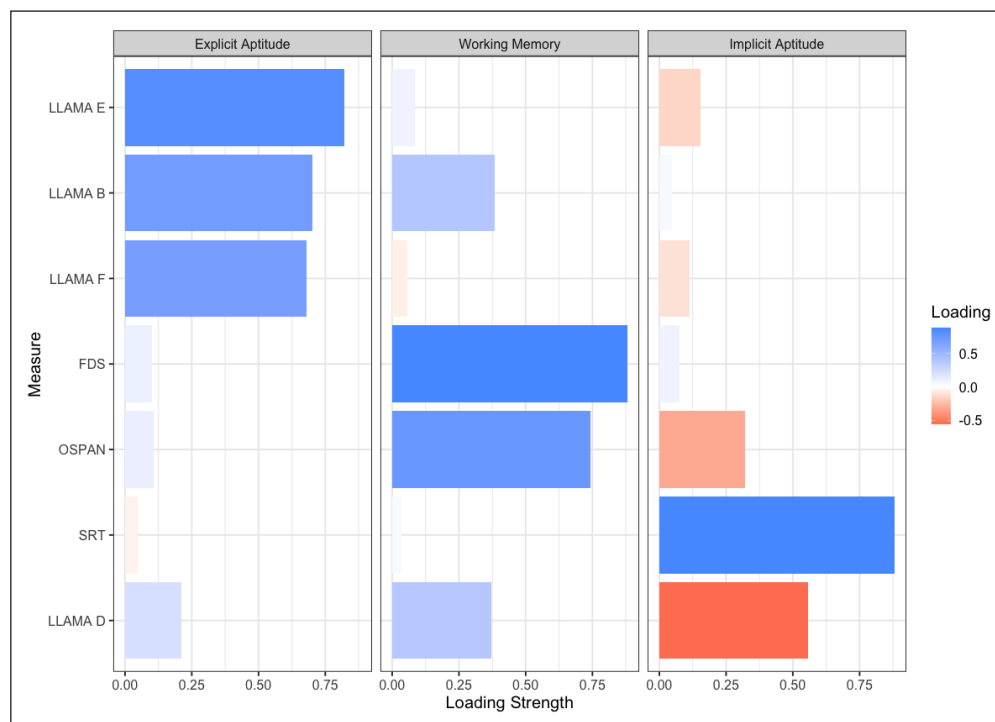
The SRT task and LLAMA D, the two hypothesized measures of implicit aptitude, loaded on the same factor and distinct from other measures of explicit aptitude (factor 1) and WM (factor 2). While the initial factor loadings were as expected, it is noteworthy that the SRT task and LLAMA D are not correlated (see Figure 2). Furthermore, the factor loading for the SRT task is positive, while the loading for LLAMA D is negative. This discrepancy suggests a complex relationship between the two measures that cannot be fully captured with a simple convergence test.

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

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

Figure 2 shows the correlations between the measures of aptitude for explicit and implicit learning and the WM measures. LLAMA B and E, as well as LLAMA E and F, are significantly correlated at a moderate level of strength, which is in accordance with expectations. The two WM measures are likewise positively and significantly associated, as one might expect, and LLAMA B is moderately correlated with the OSPAN, suggesting an association between the ability to learn new lexical items and the central executive component of WM.

As can be seen from the high factor loadings in Figure 3, LLAMA E, B, and F as measures of explicit aptitude load on factor 1 ($\lambda = 1.97$), which accounts for 28% of the variance. The two WM measures load on factor 2 ($\lambda = 1.31$), which explains 19% of the variance. Finally, the SRT task and LLAMA D, conceptualized as measures of implicit aptitude, load on factor 3 ($\lambda = 1.12$),

which explains 16% of the variance. Taken together, the three factors explain 63% of the variance and highlight the distinct loading patterns of aptitude measures compared to those assessing WM.

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

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

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

Next, we examined the relationships between all variables, as shown in Figure 2. The three L2 proficiency measures are positively associated with each other at a medium level of strength, which is not unexpected. The measures of explicit aptitude LLAMA B, E and F are moderately but significantly correlated with reading, and LLAMA E and F are moderately correlated with the gap-fill test assessing morphosyntactic knowledge. It is worth noting that none of the hypothesized predictor variables is correlated with listening. As a consequence, we conducted two hierarchical multiple regression analyses with reading and gap-fill as dependent variables, respectively. In each analysis, predictor variables were entered in descending order according to the absolute values of their correlation coefficients with the outcome variable, as shown in Figure 2. All assumptions were met in accordance with Field (2018) and Jeon (2015). Univariate outliers (cases with a z-score larger than ± 3.3) and multivariate outliers (values with Mahalanobis distance greater than 26.125 – χ^2 [8] = 26.125, p < 0.001) were removed. Following the analysis, only variables that significantly predicted variance in the dependent variable were included in the final model.

The final model for reading as shown in Table 1 includes four predictor variables, LLAMA B, LLAMA E, SRT and LLAMA D, which accounted for 30% of the variance in reading scores. Importantly, LLAMA B and E positively predict scores in reading, while the SRT task and LLAMA D are negative predictors.

	MODEL 1		MODEL 2		MODEL 3		MODEL 4	
	B	β	B	β	B	β	B	β
Constant	76.186**		75.524**		79.176**		83.285**	
LLAMA B	0.175*	0.345	0.122	0.240	0.120	0.236	0.120*	0.237
LLAMA E			0.097*	0.271	0.089*	0.248	0.098*	0.273
SRT					–0.041*	–0.251	–0.042*	–0.256
LLAMA D							–0.117*	–0.234
R²	0.119		0.182		0.244		0.298	
F	8.521**		6.875**		6.555**		6.326**	
ΔR²	0.119		0.062		0.062		0.054	
DF	8.521 *		4.726*		5.023*		4.616*	

Table 1 Hierarchical multiple regression model for reading.

Note: *Significant at 0.05 level; **Significant at 0.001 level.

Table 2 shows the final model for the gap-fill test which includes two predictor variables explaining 19% of the variance in gap-fill scores: LLAMA F and forward digit span (FDS). In this model too there is both a positive predictor (LLAMA F) and a negative predictor (FDS).

	MODEL 1		MODEL 2	
	B	**β**	**B**	**β**
Constant	68.851**		80.167**	
LLAMA F	0.165*	0.379	0.180	0.413
FDS			−0.154*	−0.225
R²	0.144		0.193	
F	12.928**		9.096**	
ΔR²	0.144		0.049	
DF	12.928 **		4.651*	

Table 2 Hierarchical multiple regression model for gap-fill.
Note: *Significant at 0.05 level; **Significant at 0.001 level.

4. DISCUSSION

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

4.1. WM AS A COMPONENT OF APTITUDE

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

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

4.2. IMPLICIT APTITUDE AS A MULTI-COMPONENTIAL CONSTRUCT

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

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

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

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

These considerations arguably shed new light on the construct of aptitude more generally because they imply that more is not necessarily better. If implicit aptitude in particular were not an ability in the classic sense (i.e., higher levels are invariably advantageous), but rather

a propensity (see also Granena, 2016), where reliance on the right capacity at the right time and in the right context determines success, then this would no doubt change the outlook of L2 researchers, L2 teachers and L2 learners alike. At this point, such a line of argument is admittedly speculative. However, we believe it can usefully inform further research into the interrelations between different measures of (implicit) aptitude and their predictive power with regard to different components of L2 proficiency.

4.3. PREDICTORS OF L2 PROFICIENCY AT ADVANCED LEVELS

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

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

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

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

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

5. CONCLUSION

The present study measured explicit and implicit aptitude and WM in a group of L2 English learners of relatively advanced proficiency. Our empirical results corroborate a conceptual

differentiation between explicit and implicit aptitude on the one hand and WM on the other hand, which suggests that the use of separate measures for these constructs is advisable.

In theoretical terms, the findings that (1) the hypothesized components of implicit aptitude pulled in different directions and (2) implicit aptitude components and PWM were negative predictors of L2 grammar and reading skill encourage us to consider the possibility that (implicit) aptitude may be a cognitive proclivity rather than an ability of immutable, context-independent value. This argument is in alignment with a comment put forward by Iizuka and DeKeyser (2023, p. 17) in which they refer to aptitude considered in this way as being reminiscent of (cognitive or learning) style (see also Granena, 2016). It also chimes with earlier research taking a multi-dimensional and dynamic view of aptitude (Robinson, 2005, 2012), according to which the sensitivity of aptitude to environmental factors is such that it can be either activated or inhibited, based on the characteristics of various learning conditions. Over and above the role of learning context, our findings have highlighted the role of learners' proficiency level in the aptitude-outcome equation.

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

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

ETHICS AND CONSENT

Ethical approval for the study was obtained from the Faculty of Social Sciences ethics sub-committee at our institution prior to the start of participant recruitment and data collection.

ACKNOWLEDGEMENTS

AUTHOR AFFILIATIONS

Renato Pavlekovic orcid.org/0000-0002-5955-0379
University of Essex, United Kingdom

Karen Roehr-Brackin orcid.org/0000-0002-1152-5852
University of Essex, United Kingdom

REFERENCES

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

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

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

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

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

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

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

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

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

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

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

DeKeyser, R. M. (2021). *Aptitude-treatment interaction in second language learning.* John Benjamins. DOI: https://doi.org/10.1075/bct.116

Field, A. P. (2018). *Discovering statistics using IBM SPSS statistics* (5th ed.). SAGE.

Gevins, A., Cutillo, B., & **Smith, M. E.** (1995). Regional modulation of high resolution evoked potentials during verbal and non-verbal matching tasks. *Electroencephalography and Clinical Neurophysiology*, *94*(2), 129–147. DOI: https://doi.org/10.1016/0013-4694(94)00261-I

Godfroid, A., & **Kim, K. M.** (2021). The contributions of implicit-statistical learning aptitude to implicit second-language knowledge. *Studies in Second Language Acquisition*, *43*(3), 606–634. DOI: https://doi.org/10.1017/S0272263121000085

Granena, G. (2012). *Age differences and cognitive aptitudes for implicit and explicit learning in ultimate second language attainment*. University of Maryland, College Park.

Granena, G. (2013a). Cognitive aptitudes for second language learning and the LLAMA Language Aptitude Test. In G. Granena & M. H. Long (Eds.), *Sensitive periods, language aptitude, and ultimate L2 attainment* (pp. 105–130). John Benjamins. DOI: https://doi.org/10.1075/lllt.35.04gra

Granena, G. (2013b). Individual differences in sequence learning ability and second language acquisition in early childhood and adulthood. *Language Learning*, *63*(4), 665–703. DOI: https://doi.org/10.1111/lang.12018

Granena, G. (2016). Cognitive aptitudes for implicit and explicit learning and information-processing styles: An individual differences study. *Applied Psycholinguistics*, *37*(3), 577–600. DOI: https://doi.org/10.1017/S0142716415000120

Granena, G. (2019). Cognitive aptitudes and L2 speaking proficiency: Links between LLAMA and Hi-LAB. *Studies in Second Language Acquisition*, *41*(2), 313–336. DOI: https://doi.org/10.1017/S0272263118000256

Granena, G. (2020). *Implicit language aptitude*. Cambridge University Press. DOI: https://doi.org/10.1017/9781108625616

Granena, G., & **Yilmaz, Y.** (2019). Corrective feedback and the role of implicit sequence-learning ability in L2 online performance. *Language Learning*, *69*(S1), 127–156. DOI: https://doi.org/10.1111/lang.12319

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

IBM Corp. (2020). *IBM SPSS Statistics for Macintosh* (Version 27.0). IBM Corp.

Iizuka, T., & **DeKeyser, R.** (2023). Scrutinizing LLAMA D as a measure of implicit learning aptitude. *Studies in Second Language Acquisition.* Advance online publication. DOI: https://doi.org/10.1017/S0272263122000559

Jeon, E. H. (2015). Multiple regression. In L. Plonsky (Ed.), *Advancing quantitative methods in second language research* (pp. 131–158). Routledge. DOI: https://doi.org/10.4324/9781315870908-7

Jiménez, L., & **Vázquez, G. A.** (2005). Sequence learning under dual-task conditions: Alternatives to a resource-based account. *Psychological Research*, *69*(5), 352–368. DOI: https://doi.org/10.1007/s00426-004-0210-9

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

Kaufman, S. B., DeYoung, C. G., Gray, J. R., Jiménez, L., Brown, J., & Mackintosh, N. (2010). Implicit learning as an ability. *Cognition*, 116(3), 321–340. DOI: https://doi.org/10.1016/j.cognition.2010.05.011

Li, S. (2013). The interactions between the effects of implicit and explicit feedback and individual differences in language analytic ability and working memory. *The Modern Language Journal*, 97(3), 634–654. DOI: https://doi.org/10.1111/j.1540-4781.2013.12030.x

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

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

Li, S. (2022). Explicit and implicit language aptitudes. In S. Li, P. Hiver, & M. Papi (Eds.), *The Routledge handbook of second language acquisition and individual differences* (pp. 37–53). Routledge. DOI: https://doi.org/10.4324/9781003270546

Li, S., & DeKeyser, R. (2021). Implicit language aptitude: Conceptualizing the construct, validating the measures, and examining the evidence: Introduction to the special issue. *Studies in Second Language Acquisition*, 43(3), 473–497. DOI: https://doi.org/10.1017/S0272263121000024

Li, S., & Qian, J. (2021). Exploring syntactic priming as a measure of implicit language aptitude. *Studies in Second Language Acquisition*, 43(3), 574–605. DOI: https://doi.org/10.1017/S0272263120000698

Linck, J. A., Hughes, M. M., Campbell, S. G., Silbert, N. H., Tare, M., Jackson, S. R., Smith, B. K., Bunting, M. F., & Doughty, C. J. (2013). Hi-LAB: A new measure of aptitude for high-level language proficiency. *Language Learning*, 63(3), 530–566. DOI: https://doi.org/10.1111/lang.12011

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

Linck, J. A., & Weiss, D. J. (2011). Working memory predicts the acquisition of explicit L2 knowledge. In C. Sanz & R. P. Leow (Eds.), *Implicit and explicit language learning* (pp. 101–114). Georgetown University Press.

Masgoret, A. M., & Gardner, R. C. (2003). Attitudes, motivation, and second language learning: A meta-analysis of studies conducted by Gardner and associates. *Language Learning*, 53(1), 123–163. DOI: https://doi.org/10.1111/1467-9922.00212

Meara, P. (2005). *LLAMA Language Aptitude Tests: The manual*. Lognostics. DOI: https://doi.org/10.1037/t88467-000

Meara, P., & Rogers, V. (2019). *The LLAMA tests v3*. Lognostics.

Oxford University Press. (n.d.). *Oxford Placement Test*. https://elt.oup.com/feature/global/oxford-online-placement/?cc=gb&selLanguage=en

Peirce, J., Gray, J. R., Simpson, S., MacAskill, M., Höchenberger, R., Sogo, H., Kastman, E., & Lindeløv, J. K. (2019). PsychoPy2: Experiments in behavior made easy. *Behavior Research Methods*, 51(1), 195–203. DOI: https://doi.org/10.3758/s13428-018-01193-y

R Core Team. (2021). *R: A language and environment for statistical computing* (Version 2021.09.2). R Foundation for Statistical Computing. https://www.R-project.org/

Robinson, P. (2005). Aptitude and second language acquisition. *Annual Review of Applied Linguistics*, 25, 46–73. DOI: https://doi.org/10.1017/S0267190505000036

Robinson, P. (2012). Individual differences, aptitude complexes, SLA processes, and aptitude test development. In M. Pawlak (Ed.), *New perspectives on individual differences in language learning and teaching* (pp. 57–76). Springer. DOI: https://doi.org/10.1007/978-3-642-20850-8_4

Roehr, K., & Gánem-Gutiérrez, G. A. (2009). The status of metalinguistic knowledge in instructed adult L2 learning. *Language Awareness*, 18(2), 165–181. DOI: https://doi.org/10.1080/09658410902855854

Roehr-Brackin, K., Gánem-Gutiérrez, G. A., Olivera-Smith, L., & Torres-Marín, M. T. (2021). Are individual differences in cognitive abilities and stylistic preferences related to multilingual adults' performance in explicit learning conditions? *Language Awareness*, 30(4), 391–412. DOI: https://doi.org/10.1080/09658416.2021.1969404

Roehr-Brackin, K., Loaiza, V., & Pavlekovic, R. (2023). Language learning aptitude in older adults. *Journal of the European Second Language Association*. DOI: https://doi.org/10.22599/jesla.93

Rogers, V. E., Meara, P., Aspinall, R., Fallon, L., Goss, T., Keey, E., & Thomas, R. (2016). Testing aptitude: Investigating Meara's (2005) LLAMA tests. *EUROSLA Yearbook*, 16(1), 179–210. 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*. DOI: https://doi.org/10.22599/jesla.24

Sáfár, A., & Kormos, J. (2008). Revisiting problems with foreign language aptitude. *International Review of Applied Linguistics in Language Teaching*, 46(2), 113–136. DOI: https://doi.org/10.1515/IRAL.2008.005

Pavlekovic and Roehr-   **17**
Brackin
*Journal of the European
Second Language
Association*
DOI: 10.22599/jesla.117

Saito, K., Sun, H. U. I., & **Tierney, A.** (2019). Explicit and implicit aptitude effects on second language speech learning: Scrutinizing segmental and suprasegmental sensitivity and performance via behavioural and neurophysiological measures. *Bilingualism: Language and Cognition*, 22(5), 1123–1140. DOI: https://doi.org/10.1017/S1366728918000895

Serafini, E. J., & **Sanz, C.** (2016). Evidence for the decreasing impact of cognitive ability on second language development as proficiency increases. *Studies in Second Language Acquisition*, 38(4), 607–646. DOI: https://doi.org/10.1017/S0272263115000327

Skehan, P. (1998). *A cognitive approach to language learning.* Oxford University Press. DOI: https://doi.org/10.1177/003368829802900209

Skehan, P. (2002). Theorising and updating aptitude. In P. Robinson (Ed.), *Individual differences and instructed language learning* (pp. 69–94). John Benjamins. DOI: https://doi.org/10.1075/lllt.2.06ske

Skehan, P. (2016). Foreign language aptitude, acquisitional sequences, and psycholinguistic processes. In G. Granena, D. O. Jackson, & Y. Yilmaz (Eds.), *Cognitive individual differences in second language processing and acquisition* (pp. 17–40). John Benjamins. DOI: https://doi.org/10.1075/bpa.3.02ske

Suzuki, Y., & **DeKeyser, R.** (2015). Comparing elicited imitation and word monitoring as measures of implicit knowledge. *Language Learning*, 65(4), 860–895. DOI: https://doi.org/10.1111/lang.12138

Suzuki, Y., & **DeKeyser, R.** (2017). The interface of explicit and implicit knowledge in a second language: Insights from individual differences in cognitive aptitudes. *Language Learning*, 67(4), 747–790. DOI: https://doi.org/10.1111/lang.12241

Tabachnick, B. G., & **Fidell, L. S.** (2014). *Using multivariate statistics* (6th ed.). Pearson.

Teimouri, Y., Goetze, J., & **Plonsky, L.** (2019). Second language anxiety and achievement: A meta-analysis. *Studies in Second Language Acquisition*, 41(2), 363–387. DOI: https://doi.org/10.1017/S0272263118000311

Unsworth, N., Heitz, R. P., Schrock, J. C., & **Engle, R. W.** (2005). An automated version of the operation span task. *Behavior Research Methods*, 37(3), 498–505. DOI: https://doi.org/10.3758/BF03192720

Wen, Z. E. (2019). Working memory as language aptitude: The Phonological/Excecutive Model. In Z. E. Wen, P. Skehan, A. Biedrón, S. Li, & R. L. Sparks (Eds.), *Language aptitude: Advancing theory, testing, research and practice* (pp. 277–338). Routledge. DOI: https://doi.org/10.4324/9781315122021

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.** (2019). Cognitive individual differences as predictors of improvement and awareness under implicit and explicit feedback conditions. *The Modern Language Journal*, 103(3), 686–702. DOI: https://doi.org/10.1111/modl.12587

Yi, W. (2018). Statistical sensitivity, cognitive aptitudes, and processing of collocations. *Studies in Second Language Acquisition*, 40(4), 831–856. DOI: https://doi.org/10.1017/S0272263118000141

Yoshimura, Y. (2001). The role of working memory in language aptitude. In X. Bonch- Bruevich, W. J. Crawford, J. Hellermann, C. Higgins, & H. Nguyen (Eds.), *The past, present, and future of second language research* (pp. 144–163). Cascadilla Press.