Cognitive Aptitudes and Processing of L2 Grammar: Exploring the Role of Rule Inferencing Skills and Working Memory

Małgorzata Foryś-Nogala

Faculty of Psychology, University of Economics and Human Sciences in Warsaw
Faculty of Psychology, University of Warsaw

ABSTRACT

This correlational study investigated the relationship between cognitive aptitudes and online and offline processing of L2 syntactic structures. As a measure of online processing, the study used a self-paced reading task. To tap into offline L2 knowledge, it employed an untimed grammaticality judgment task (GJT). The main analyses focused on the correct placement of relative pronouns. The supplementary analyses were carried out on a range of other structures used as fillers in the GJT. In terms of cognitive aptitudes, the study considered the role of explicit learning aptitude and working memory in the processing of L2. Explicit aptitude was operationalized as an ability to infer rules of a new language and measured by the LLAMA F task, and working memory was measured by a digit span task. Moreover, the design included a measure of general L2 proficiency. The results showed that L2 learners' scores in the GJT were positively related to their explicit language aptitude. However, this type of relationship was observed only for ungrammatical items. In contrast, working memory was not a significant predictor of the performance on the GJT. As regards online processing, no links were found between the predictor variables and participants' sensitivity to errors in the self-paced reading task. Taken together, the results corroborate the role of explicit learning abilities in offline processing of L2 grammar. Additionally, supplementary analyses suggest that this relationship may hold even when general L2 proficiency is controlled for.

KEYWORDS

implicit L2 knowledge
explicit L2 knowledge explicit
learning aptitude
rule inferencing skills
working memory

INTRODUCTION

Second language acquisition (SLA) research pursues to explain why people differ in their success rates in learning foreign/ second languages in adulthood. Factors proposed to determine L2 development encompass the context of L2 acquisition (e.g., naturalistic acquisition versus instructed learning), instructional treatments (i.e., a method of instruction), and features of the target language (e.g., its typological proximity to L1), as well as learner-related variables, including emotions, motivations, and cognitive aptitudes (e.g., DeKeyser, 2012; Granena & Long, 2013; Wen et al., 2017; Yilmaz & Granena, 2021). Cognitive language aptitude is a relatively stable characteristic of an individual that determines L2 learning success when other factors (e.g., the context of instruction, motivation) are held constant (Doughty, 2019). Traditionally, the explicit component of language aptitude was the main point of focus in SLA studies (see Yilmaz & Granena, 2021, for a discussion). However, due to a growing body of research on implicit L2 knowledge, implicit learning aptitude has attracted more and more attention of SLA scholars (e.g., Granena, 2013; Linck et al., 2013; Suzuki & DeKeyser, 2017). Likewise, accounting for predictors of both explicit and implicit L2 processing was brought into the spotlight in research on L2 acquisition. In the words of Suzuki and DeKeyser (2017), "explicit/implicit learning processes can be inferred from the relationship between knowledge (i.e., product of learning) and aptitude" (p. 753). Nevertheless, there are relatively few studies simultaneously exploring cognitive predictors of both explicit and implicit knowledge of L2 that use fine-grained psycholinguistic methods to measure implicit knowledge (see Suzuki, 2017 for a methodological discussion). Moreover, existing evidence on the links between cognitive aptitudes and L2 knolwedge is mixed. In particular, it is not settled how work-

Corresponding author: Małgorzata Foryś-Nogala, University of Economics and Human Sciences in Warsaw, ul. Okopowa 59, 01-043 Warsaw, Poland. E-mail: m.forys-nogala@vizja.pl

DOI • 10.5709/acp-0340-1

ing memory capacity influences online and offline processing of L2. Another underexplored issue involves the nature of links between implicit knowledge and explicit learning aptitude. Therefore, the current study investigated the relationships between individual differences in cognitive aptitudes– rule inferencing skills and working memory– and online (explicit) and offline (implicit) processing of L2 syntactic structures in a sample of adult Polish learners of L2 English. Offline processing, assumed to reflect explicit L2 knowledge, was measured by an untimed grammaticality judgment task (GJT) and online processing, assumed to reflect implicit L2 knowledge, was tapped into by a self-paced reading (SPR) task.

LITERATURE BACKGROUND

In SLA research, language learning aptitudes are often considered in relation to the distinction between implicit and explicit knowledge of L2 (e.g., DeKeyser, 2012; Granena, 2013; Linck et al., 2014; Pawlak & Biedroń, 2021; Suzuki & DeKeyser, 2017). Implicit knowledge has been defined as knowledge without awareness, which does not rely on conscious metalinguistic rules (Hulstijn, 2002; Paradis, 2009). Methodological research has shown that this knowledge can be best captured by fine-grained psycholinguistic measures of online processing such as a word monitoring task, an SPR task, and the visual world paradigm (Suzuki, 2017; Vafaee et al., 2017). In contrast, explicit knowledge is underlain by conscious processes and verbalizable metalinguistic rules (e.g., Williams, 2009). Explicit knowledge is typically measured by untimed focus-on-form tasks, such as the GJT. Crucially, advanced L2 users typically develop a type of explicit knowledge called *automatized explicit knowledge*, whose processing is fast yet still reliant on consciously accessed metalinguistic rules (e.g., Hulstijn, 2002). Automatized explicit knowledge is tapped into by timed focus-on-form tasks, such as timed GJTs or elicited imitation tasks (e.g., Suzuki & DeKeyser, 2015, 2017). Research into language learning aptitudes has underscored that explicit and implicit language aptitudes may differentially affect the growth of implicit and explicit knowledge (e.g., Granena, 2013; Yilmaz & Granena, 2021). However, the pattern of relationships between knowledge types and aptitudes is still largely unexplored.

Explicit language aptitude involves the ability to derive rules from language samples through logical reasoning, identifying patterns in the input and testing hypotheses (e.g., Yilmaz & Granena, 2021). Skehan (1998) called the construct 'language analytic ability' (LAA) and defined it as "the capacity to infer rules of language and make linguistic generalization or extrapolations" (p. 204). Previous research showed positive relationships between explicit learning aptitude, measured by tasks involving novel languages, and scores in offline untimed measures of grammatical knowledge (Abrahamsson & Hyltenstam, 2008; DeKeyser, 2000). DeKeyser (2000) tested a sample of L2 learners diversified in terms of L2 proficiency as well as age of arrival in the country of L2 and found correlations between verbal analytical ability and scores on the GJT. However, this link was reported for adult arrivals but not for child arrivals. In contrast, Abrahamsson and Hyltenstam (2008) focused on near-native L2 speakers and found that both late and early L2 learners benefited from high language aptitude (but the effect was smaller in the case of early arrivals). Moreover, experimental studies involving explicit instruction (e.g., including explicit corrective feedback or deductive learning methods) found that L2 learning is facilitated by explicit language aptitude (e.g., De Graaff, 1997; Yilmaz, 2013). As for automatized explicit knowledge, in their correlational cross-sectional study, Suzuki and DeKeyser (2017) found the link between grammar inductive ability, measured by the LLAMA F task (Meara, 2005), and measures of automatized explicit knowledge in a group of 100 advanced learners of L2 Japanese. However, Granena and Long (2013) did not show links between scores in explicit learning ability and a time-pressured GJT in their study on Chinese learners of L2 Spanish, regardless of their onset of L2 acquisition.

In contrast, implicit learning aptitude concerns the ability to extract patterns from a high-frequency input in an automatic and unconscious fashion (e.g., Granena, 2013). Many studies found correlations between implicit learning aptitude, typically measured via a serial reaction time (SRT) task (e.g., Kaufman et al., 2010), and online language processing (e.g., Granena, 2013; Suzuki & DeKeyser, 2015; Faretta-Stutenberg & Morgan-Short, 2018, in their study-abroad condition). Still, there are also studies in which this relationship was not evidenced (Suzuki & DeKeyser, 2017; Faretta-Stutenberg & Morgan-Short, 2018, in the case of at-home learners). For instance, Suzuki and DeKeyser (2017) did not report any links between implicit learning mechanisms and performance on the L2 tasks tapping into either automatized explicit or implicit knowledge. However, as mentioned above, they did find a correlation between explicit learning aptitude and automatized L2 knowledge, which was, in turn, related to implicit knowledge of the same set of structures. The authors suggested that explicit aptitudes are far more advantageous in the context of form-focused instruction than implicit aptitudes. The results obtained by Faretta-Stutenberg and Morgan-Short (2018), who investigated L2 development in two exposure conditions (immersion vs. classroom instruction), largely corroborate this scenario. Their study found links between both online (measured via ERPs[1]) and offline (measured with a GJT) L2 processing and performance on implicit learning ability tasks. However, those relationships held only when L2 learning took place via immersion, and not when it was acquired in the classroom setting (at the students' home university).

Another construct of interest in language learning aptitude research is the construct of working memory (WM), which is a system capable of a short-term storage and manipulation of information needed to perform the task at hand (Baddeley, 2003). Mounting evidence suggests that WM is a factor accounting for variability in L2 acquisition (Juffs & Harrington, 2011; Linck et al., 2014; Wen, 2016). Working memory might be implicated in L2 learning in both immersion and classroom focus-on-form settings. In the context of immersion, WM may facilitate the processing of large quantities of input during real-life communication in L2, whereas in the L2 classroom, it may enhance the learning and application of metalinguistic rules. For instance, in their correlational study on advanced learners of L2 English, Erçetin and Alptekin (2013) found links between performance on GJTs (both timed and untimed) as well as an oral imitation task and scores on a reading span task (Daneman & Carpenter, 1980), which was administered in L2 English. In a different correlational study, Pawlak and Biedroń (2021) investigated the links between WM ability and knolwedge of L2 grammar in a sample of university students majoring in English. They found positive but weak relationships between components of WM and tasks tapping into productive knowledge of L2 grammatical structures. However, no such correlations were found between receptive grammatical knowledge and memory capacity measures used in the study. Moreover, Faretta-Stutenberg and Morgan-Short (2018) did not find any correlations between WM ability and scores in an offline grammaticality judgment task in the at-home condition of their study. Similarly, automatized explicit knowledge of L2 grammar was not related to phonological short-term memory in the study by Suzuki and DeKeyser (2017). Admittedly, few studies considered the role of WM components in online L2 processing. For instance, Faretta-Stutenberg and Morgan-Short (2018) showed a relationship between WM capacity and online language processing, measured by ERPs, in the case of learners acquiring L2 in the study-abroad (i.e., immersion) condition. However, in the study by Suzuki and DeKeyser (2017), the composite score in tasks tapping into implicit L2 knowledge was not related to phonological short-term memory.

Taken together, the results of studies discussed above point to a relationship between explicit language aptitude, typically measured with inductive grammar learning tasks, and explicit knowledge of L2 grammar. This link might extrapolate to automatized explicit knowledge, however, the results are less consistent in this respect. In contrast, explicit learning aptitude is less likely to be directly linked to implicit L2 knowledge. Evidence regarding the role of WM in predicting L2 grammatical knowledge (both online and offline) is mixed. While some studies found links between WM measures and L2 grammar, some suggested that WM might not be very prominent in predicting knowledge of specific structures in L2. Moreover, many of the existing studies looked at immersion learners with a respectable length of residence in the country of L2 and relatively few studies on less proficient learners used fine-grained tasks tapping into implicit knowledge (as recommended by Suzuki, 2017). Thus, further research with more diversified samples of L2 learners is needed to explore the role of cognitive aptitudes in implicit and explicit L2 processing.

THE PRESENT STUDY

This study aimed to investigate the predictive relationships between language learning aptitude and processing of L2 grammatical structures. Specifically, it focused on the role of WM and explicit learning aptitude in acquiring selected features of L2 English by Polish learners while controlling for their general L2 proficiency. The study pursued the following research questions:

1. Is explicit learning aptitude differentially related to explicit and implicit knowledge of L2 grammatical structures when L2 proficiency is controlled for?
2. Is a composite score in WM capacity a predictor of offline and online processing of L2 grammar?

METHOD

Participants

The participants were 69 adult native speakers of Polish (48 females; M_{age} = 22.65; SD = 4.77), who completed all experimental tasks meant for the present analysis. They were recruited through a job search website for a larger project on incidental vocabulary learning. Data from one participant were excluded from the analysis due to a very unusual pattern of responses in the GJT, suggesting that she had not followed the instructions. Of the 68 participants, 19 were university graduates, 42 were high school graduates, and the remaining 7 participants were high school undergraduates. Ten participants reported having lived in a foreign country for longer than three months; however, only in the case of two participants, the country of residence was an English-speaking country (their length of residence in that country was 4 and 6 months, respectively). Apart from knowledge of English as a foreign language, 56 participants reported knowledge of other foreign languages. All participants were provided with the general information about the study's aim and procedure and were informed about their unconditional right to withdraw at any stage of the study. They all gave their written informed consent to participate and received financial renumeration for their contribution.

Materials

TARGET STRUCTURE

The study focused on the contrast in the use of relative pronouns "which" and "what" in chunks, such as: (1) "This is what…" versus (2) "This is the thing which…". According to generative syntax, "what" in (1) introduces a complementizer phrase, while "which" in (2) is a part of a determiner phrase. Earlier research found that this structure might be problematic for Polish learners of L2 English (Foryś-Nogala, 2019; Gozdawa-Gołębiowski & Opacki, 2018). For instance, in their corpus study, Gozdawa-Gołębiowski and Opacki (2018) reported that learners frequently misused "what" as a complementizer in such expressions as "*Something what is not very interesting" (p. 142). Apart from the main structure, the offline task used in the present study incorporated filler structures including aspects of subject-verb agreement, double negation, bare infinitives, and subject placement. For the online and offline L2 tasks, the study used a subset of linguistic items developed and validated for Foryś-Nogala (2019).

PROFICIENCY TEST

Participants' general proficiency in L2 English was measured by the Cambridge placement test for schools (https://www.cambridgeenglish.org/test-your-english/for-schools/; maximum score: 25). The test taps into integrated knowledge of L2 English manifested through accurate responses to a range of conversational situations.

GRAMMATICALITY JUDGMENT TASK

As a measure of offline language processing tapping into explicit L2 knowledge, the study used an untimed GJT. The task included 24 experimental items with correct or erroneous instantiations of relative pronouns "which" (12 items) and "what" (12 items). Half of the items were ungrammatical (i.e., the target relative pronoun was replaced with its alternative, thus rendering the item incorrect, see Table 1).

On each trial, together with a written stimulus, the participants saw a picture showing an object or person described by a given sentence. The grammaticality of the experimental sentences was manipulated within-item. Therefore, two versions of the GJT, differing in the grammaticality status of each of the target items, were counterbalanced across the participants. Apart from experimental items, the task included 36 filler items, half of which were ungrammatical. Ungrammatical filler items contained violations of such structures as subject-verb agreement, bare infinitives, negation, and subject placement (see Table 1). A priori, fillers were incorporated to mask the prominence of the target structure, therefore, they were not manipulated within-item. However, they were also used in the supplementary analyses of the GJT scores (see the Results section). Experimental items in both versions of GJT displayed satisfactory reliability of Cronbach's α = 0.73 and α = 0.77. However, for the filler items, which contained a variety of L2 structures, the coefficient was lower, α = 0.65.

SELF-PACED READING

To investigate online language processing, the study used a moving window SPR task (Just et al., 1982). There were 14 items per each target structure (28 experimental items in total). In half of the experimental items, "which" was replaced with "what" and vice versa, rendering them ungrammatical (see Table 2). Experimental trials were made of pairs of similar sentences, which differed in the picture stimuli, comprehension question and a minor detail (e.g., a different temporal adverb). There were also 36 grammatical fillers and 5 training items in the task.

TABLE 1.
Examples of Experimental and Filler Items in the Untimed Grammaticality Judgment Task

Structure status	Structure type	Example item
experimental	relative pronoun (what)	Have a look… This is what/*which Jake sent to me from his vacation in Madrid.
experimental	relative pronoun (which)	Have a look.. This is the picture which/*what someone drew in my history book.
filler	SV agreement	Have a look… This is my neighbour, John. He is so friendly that everybody *like him.
filler	double negation	Have a look… There aren't *no oranges in the basket.
filler	subject placement	Have a look… A lot has changed in John's life since *started a family.
filler	bare infinitives	My daughter has such bad grades that she will *has to take additional classes.

Note. Asterisks mark grammatical errors.

Each trial consisted of 11 words and unfolded upon a button press. The presentation of each item commenced with a short introductory phrase, followed by a picture, and finished with a control "Yes/No" question. The order of the presentation of items was randomized. As in the GJT, the study used two parallel versions of the SPR task (differing in the grammaticality status of experimental items), which were counterbalanced across the participants. As indicated by Cronbach's α , both versions of the task reached good reliability of α = .87 and α = .93, for "what" items, and α = .86 and α = .89 for "which" items[2]. Examples of experimental and filler items are presented in Table 2

LLAMA F – A RULE INFERENCING TASK

As a measure of explicit learning aptitude, the study used the LLAMA F task from the LLAMA aptitude test battery (Meara, 2005). The LLAMA F is a computerized task tapping into rule inferencing skills, in which the participants are asked to analyze phrases in an unknown language shown together with pictures illustrating their meaning. The training lasts for 5 minutes and the subsequent testing phase is untimed. In the test, participants are asked to choose one of two sentences that best describes a picture shown in the task panel. The raw scores in LLAMA F range between 0 and 100. The present study used the default task administration procedure.

DIGIT SPAN

As s a measure of working memory, the study used the Digit Span task from the Polish adaptation of the Wechsler Adult Intelligence Scale (WAIS-R [PL], Brzeziński et al., 2004), which was administered individually in a quiet room. The task involves forward (Part 1) and backward (Part 2) repetition of strigs of digits. The total raw score in the task was 28 points (14 points per block). For the present analysis, the study used a composite score in both blocks of the task to capture both storage and processing components of WM.

Procedure

The present study was embedded within a larger project on incidental vocabulary learning, whose procedure extended over five days (Silva et al., in preparation). Notably, the project involved a two-stage prescreening procedure to limit floor or ceiling effects in the vocabulary

TABLE 2.
Examples of Experimental and Filler Items in the Self-Paced Reading Task

Structure status	Structure type	Example item
experimental	relative pronoun (what)	Have a look at this picture! This/ is/ **what(*which)**/ Emma/ left/ at/ her/ friend's/ house/ last/ week.
experimental	relative pronoun (which)	Have a look at this picture. This/ is/ the/ desk/ **which(*what)**/ Ben/ made/ for/ my/ younger/ sister.
filler	-	Oh no! There/ isn't/ any/ sugar/ in/ the/ only/ shop/ in/ our/ village.

Note. Critical segments are given in bold.

tasks, following which 72 participants were qualified for the main study. During the main study sessions, the participants completed vocabulary tasks as well as additional cognitive and linguistic measures, including the tasks whose results are reported here. As for those, Day 1 involved the Cambridge placement test for schools; Day 2 involved the LLAMA F (Meara, 2005); Day 3 involved the SPR task; Day 4 involved the grammaticality judgment task; and, finally, Day 5 involved the WM task. The GJT and the SPR task were administered via PsychoPy (Peirce et al., 2019). The procedure of the present study was approved by the Research Ethics Committee of the Faculty of Psychology, University of Warsaw. The procedure of the larger project within which this study was embedded was also approved by the Rector's Ethics Committee of the University of Warsaw (application no. 9/2017).

RESULTS

Subject-Related Variables

Table 3 shows descriptive statistics for the measures tapping into general L2 proficiency, WM and inductive grammar learning ability measured by the LLAMA F, as well as descriptive statistics for the offline GJT. As regards L2 proficiency, the scores in the Cambridge placement test for schools showed that there was considerable variability in terms of L2 knowledge among the participants (see Table 3). The average raw score in the test was 19.68 points out of 25 (79%), which corresponds to level B1/B2 according to the thresholds retrieved from the test website. Overall, items in the GJT were relatively easy for the participants, with grammatical experimental items reaching ceiling scores. Since ungrammatical items reached lower raw scores than grammatical items for both experimental and filler structures, grammaticality was introduced as a fixed effect in the linear mixed models presented below.

Offline Grammaticality Judgment Task

Scores on the offline GJT were analyzed with mixed-effects logistic regression models fitted using the *glmer* function in the *lme4* package (Bates et al., 2015) in *R* (R Core Team, 2020) with the Bound Optimization by Quadratic Approximation (BOBYQA) optimizer (Powell, 2009). The *p* values in the models were estimated using the *lmerTest* package (Kuznetsova et al., 2017). Initially, the models adopted the maximal random effects structure (Barr et al., 2013), which was reduced by removing random slopes equal (or very close) to 0. All continuous predictors were centred and all categorical predictors were deviation-coded. The intercepts of the models reflect the grand mean.

In the model computed for the experimental items, the dependent variable was the accuracy of assessing each sentence as either grammatical or ungrammatical. The fixed effects included: grammaticality of the item (ungrammatical versus grammatical), type of the target pronoun ("which" versus "what"), L2 proficiency, score in the LLAMA F, a composite score in the Digit Span task, as well as the interactions between grammaticality and pronoun type and grammaticality and participant-related variables, namely, the LLAMA F, L2 proficiency, and the Digit Span task. The model for experimental items (i.e., wh-pronouns) is shown in Table 4.

As shown in Table 4 (see also Table 5), the only significant predictors in the model were grammaticality and the grammaticality × LLAMA F interaction. In general, with all other fixed effects assuming mean values, grammatical sentences were about 11% more likely to be accurately assessed by the participants than ungrammatical sentences. However, this effect was reduced with increasing scores in inductive grammar learning ability, as measured by the LLAMA F. As depicted in Figure 1, chances for the correct assessment of ungrammatical items increased with higher scores in the LLAMA F. In contrast, for grammatical items, this relationship was reversed. Importantly, since the performance on correct items was at ceiling levels, the latter observation needs to be approached with caution.

Next, a supplementary analysis was ran on the filler items to see if the pattern replicates on a range of different and more diversified sentence features. As explained in the method, the filler items' grammaticality was not manipulated within-item. Therefore, the by-item random slope for grammaticality was not included in the analysis. Moreover, as fillers were not systematically grouped into structure types, it was impossible to include the type of structure as a fixed effect in the model. As shown in Table 6 (see also Table 7), for the filler items, the likelihood of a correct answer in the GJT increased with ascending proficiency, however, this effect was not moderated by grammaticality. Also, as in the case of the experimental items, grammatical filler items stood a bigger chance of being assessed correctly than ungrammatical items (by about 9%). Notably, the interaction between grammaticality and the LLAMA F

TABLE 3.
Descriptive Statistics for the Aptitude and Linguistic Tasks

Measure	*M*	*SD*	Min-Max
L2 proficiency	0.79	0.11	52–96%
Digit Span (max. 28)	14.66	3.55	7–24
LLAMA F	0.62	0.28	0–100%
GJT – all wh-pronouns	0.83	0.14	46–100%
GJT – ungrammatical wh-pronouns	0.75	0.25	0–100%
GJT – grammatical wh-pronouns	0.92	0.10	67–100%
GJT – ungrammatical filler items	0.77	0.16	28–100%
GJT – grammatical filler items	0.86	0.11	50–100%

TABLE 4.
Fixed Effects Estimates in the Model for Experimental Items

	Estimate	*SE*	*z*	*p*
Intercept	2.47	0.20	12.58	< .001***
Grammaticality (ungrammatical vs. grammatical)	1.53	0.31	4.96	< .001***
Pronoun type ("which" vs. "what")	0.20	0.25	0.82	.411
Proficiency	2.17	1.54	1.41	.158
LLAMA F	0.24	0.59	0.40	.686
Digit Span	0.03	0.05	0.74	.462
Grammaticality × Pronoun type	−0.70	0.47	−1.49	.136
Grammaticality × LLAMA F	−3.01	1.01	−2.98	.003**
Grammaticality × Proficiency	−1.51	2.59	−0.58	.561
Grammaticality × Digit Span	0.07	0.08	0.81	.418

Note. Number of data points: 1632; Subjects: 68; Items: 24

TABLE 5.
Random Effects Estimates in the Model for Experimental Items

Groups	Effect	Variance	*SD*
Subject	(Intercept)	1.15	1.07
	Pronoun type	1.50	1.23
	Grammaticality	2.69	1.64
	Grammaticality × Pronoun type	1.97	1.40
Item	(Intercept)	0.02	0.15
	Grammaticality	0.32	0.56

Note. Number of data points: 1632, Subjects: 68; Items: 24

TABLE 6.
Fixed Effects Estimates in the Model for Filler Items

	Estimate	*SE*	*z*	*p*
Intercept	1.83	0.15	11.92	$< .001^{***}$
Grammaticality (ungramatical vs. grammatical)	0.72	0.30	2.43	$.015^{*}$
Proficiency	3.16	0.84	3.76	$< .001^{***}$
LLAMA F	0.21	0.32	0.65	.519
Digit Span	0.02	0.02	0.81	.416
Grammaticality × LLAMA F	−1.10	0.62	−1.78	.075
Grammaticality × Proficiency	−0.495	1.619	−0.31	.760
Grammaticality × Digit Span	0.01	0.046	0.145	.884

Note. Number of data points: 2448; Subjects: 68; Items: 36

TABLE 7.
Random Effects Estimates in the Model for Filler Items

Groups	Effect	Variance	*SD*
Subject	(Intercept)	0.24	0.49
	Grammaticality	0.80	0.90
Item	(Intercept)	0.55	0.74
	LLAMA F	0.47	0.69
	Proficiency	3.43	1.85

Note. Number of data points: 2448; Subjects: 68; Items: 36

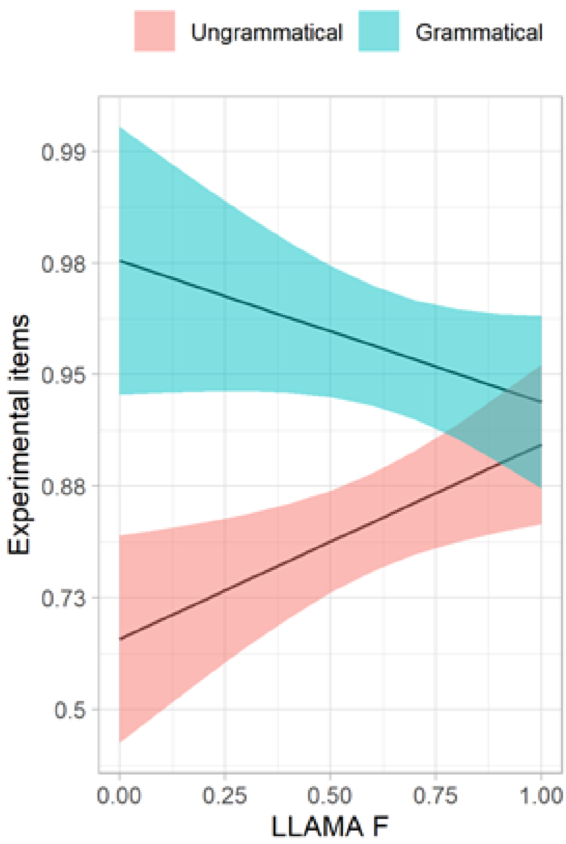

FIGURE 1.
Linear mixed effects model's estimates for the relationship between scores in the LLAMA F and responses to grammatical and ungrammatical experimental items in the offline grammaticality judgment task (GJT).

reached a trend (p = .075). As shown in Figure 2, the pattern resembled the interaction reported for the model for the experimental items.

Finally, supplementary multiple regression models were ran with aggregated scores in both experimental and filler items. The models were computed separately for ungrammatical and grammatical sentences. In both models, the predictors included centred LLAMA F scores, general proficiency in L2 English as well as the score in the Digit Span task. As for aggregated ungrammatical sentences, both LLAMA F scores (β = 0.17; SE = 0.06; t = 2.71; p = .009) and L2 proficiency (β = 0.44; SE = 0.16; t = 2.74; p = .008) were significant predictors of spotting errors in the GJT. The relationship between WM and the ungrammatical GJT was nonsignificant (β = 0.002; SE = 0.005; t = 0.47; p = .641). Overall, the model reached statistical significance, F(3, 64) = 5.83; p = .001, and accounted for 18% (corrected R^2) of the variance in the data. Importantly, scores in the LLAMA F and L2 proficiency were not intercorrelated, r = 0.06; p = .611. In the corresponding model computed for grammatical sentences, L2 proficiency was the only significant predictor of the scores (β = 0.23; SE = 0.10; t = 2.35; p = .022). The effects of the LLAMA F and WM did not reach significance (β = -0.06; SE = 0.04; t = −1.48; p = .144, and β = 0.003; SE = 0.003; t = 1.11; p = .269, respectively). The model was marginally significant, F(3, 64) = 2.90; p = .042 and accounted for 8% of the total variance in the data.

Self-Paced Reading

In the SPR task, all the participants passed the threshold of 75% score on the control questions (M = 0.90; SD = 0.05; range = 75–100%), so they were all included into the analyses (see Jiang et al., 2011). The initial data trimming involved excluding responses faster than 150 ms and slower than 3000 ms. In the next step, outliers were defined as values falling outside 3 *SD*s below or above each participant's mean (separately for grammatical and ungrammatical items) and removed from the dataset. In total, approximately 2.6% of the data points were excluded. Table 8 shows descriptive statistics for the segments of interest.

Further, raw reading times were log-transformed and, to limit the impact of individual pace of reading on the study results, residual reading times were calculated (Trueswell et al., 1994). In the first step, a series of mixed effects regression models were computed to see if the participants were sensitive to the violations of the target structures, and explore where the sensitivity emerged. For both types of structures, the analyses were carried out on the critical segment containing the target pronoun (or its ungrammatical counterpart) and two

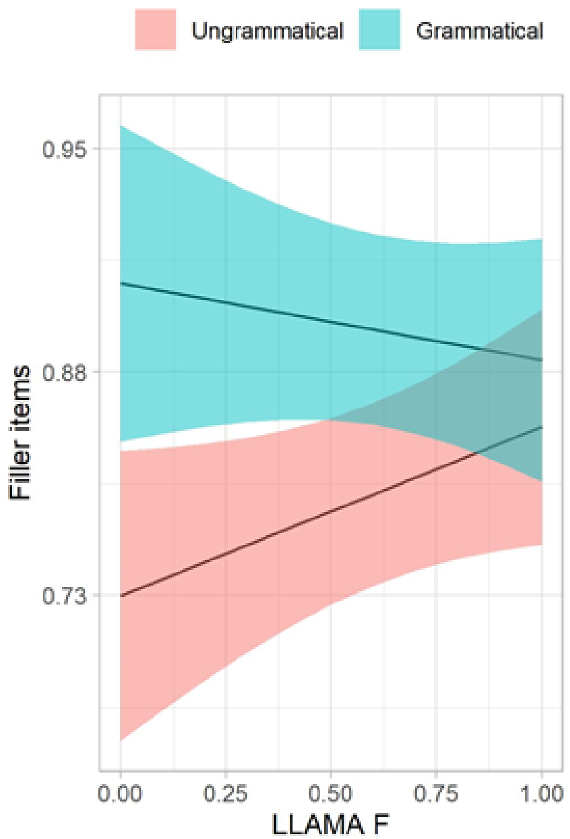

FIGURE 2.

Linear mixed effects model's estimates for the relationship between scores in LLAMA F and responses to grammatical and ungrammatical filler items in the offline grammaticality judmgent task (GJT).

subsequent segments. All the models included random intercepts for subjects and items and a by subject random slope for grammaticality. Grammaticality was deviation-coded (-0.5 denoted ungrammatical items, 0.5 denoted grammatical items). In the "what" items (i.e., where the target pronoun was "what"), the effects of reading slowdown in response to errors in pronoun placement were visible in the first ($\beta = -0.07$; SE = 0.02; $t = -3.876$; $p < .001$) and second postcritical segments ($\beta = -0.07$; SE = 0.02; $t = -3.884$; $p < .001$), but not in the critical segment ($\beta = 0.02$; SE = 0.02; $t = 1.226$; $p = .221$). Similarly, in the "which" items no grammaticality effects were observed for the critical segment ($\beta = -0.02$; SE = 0.01; $t = -1.398$; $p = .167$), but significant slowdown

TABLE 8.

Descriptive Statistics for Raw Reading Times (in Milliseconds) in the Self-Paced Reading Task

	Critical segment *M* (*SD*)	Critical segment +1 *M* (*SD*)	Critical segment +2 *M* (*SD*)
what-items			
grammatical	312.48 (85.31)	336.76 (91.64)	347.53 (92.30)
ungrammatical	313.38 (90.15)	366.58 (119.60)	376.45 (111.97)
which-items			
grammatical	335.66 (83.05)	346.65 (90.69)	368.16 (112.75)
ungrammatical	335.98 (88.28)	365.38 (114.61)	381.06 (107.87)

was noted for the first post-critical segment ($\beta = -0.04$; SE = 0.02; $t = -2.177$; $p = .030$). Also, a trend towards significance was observed for the second post-critical segment ($\beta = -0.03$; SE = 0.02; $t = -1.949$; $p = .056$). Notably, after the implementation of the Bonferroni-Holm correction for multiple comparisons, the grammaticality effects for the "which" structure were no longer statistically significant. Nevertheless, the two post-critical segments were regarded as areas where some sensitivity to the misplacement of "which" might have emerged. Consequently, they were included into further analyses. In the next step, for each post-critical segment, the sensitivity indexes were computed which reflected each participant's mean difference in reading ungrammatical and grammatical items (see Suzuki & DeKeyser, 2015, 2017). To limit the effects related to the pace reading, the indexes were computed on the residual reading times. Then, indexes related to each variant of the target structure (i.e., "which" or "what") were summed. Subsequently, multiple linear regression models were computed to explore which participant-related variables accounted for their sensitivity to the violations in the use of pronouns. Summed sensitivity indexes of the two post-critical segments were the dependent variables in the models, and the independent variables included centered scores in L2 proficiency, the Digit Span task, the LLAMA F, as well as an aggregated score in the ungrammatical experimental items in the GJT. The latter variable was included as a measure of declarative knowledge of the target contrast. As there were only six items focused on each pronoun type, scores in all ungrammatical experimental items were aggregated to check if declarative knowledge of the target structure predicted its online processing. Models for both types of pronouns failed to reach statistical significance, $F(4, 63) = 0.49$; $p = .743$, for the "what" items, and, $F(4, 63) = 0.17$; $p = .955$, for the "which" items, with none of the predictors approaching statistical significance (see Table 9).

DISCUSSION

This study explored the role of cognitive aptitudes in offline and online processing of syntactic structures in L2 English with a particular focus on explicit learning aptitude, measured by a rule inferencing task (LLAMA F) and WM, measured by the Digit Span task from the WAIS-R(PL) (the composite score in the backward and forward blocks). As an offline processing task, the study used the untimed GJT, and to tap into online (implicit) processing, it employed the moving window self-paced reading paradigm.

The results showed that explicit learning aptitude was positively related to explicit knowledge of L2 grammar. However, this relationship held only for ungrammatical sentences. With growing rule inferencing skills, the participants were more likely to identify errors in the wrong placement of relative pronouns. Further, the results of supplementary analyses carried out for the filler items, as well as for aggregated scores in both experimental and filler sentences, indicated that this relationship may generalize to a wider array of L2 features, even when some variance in the data is explained by L2 proficiency. Crucially, the positive relationship between LLAMA F and the GJT scores was reported only for erroneous items. As for grammatical sentences, some of the obtained

TABLE 9.
Coefficients in the Models Testing Predictors of Residual Reading Times in the Self-Paced Reading Task

	Estimate	*SE*	*t*	*p*
what-items				
intercept	0.14	0.03	4.87	< .001
LLAMA F	-6.52	10.58	-0.62	0.540
Proficiency	-8.39	6.73	-1.25	0.217
Digit Span	0.002	0.01	0.23	0.820
GJT (ungrammatical)	0.02	0.12	0.18	0.857
which-items				
intercept	0.07	0.03	2.46	0.017
LLAMA F	-1.34	10.44	-0.13	0.898
Proficiency	3.23	6.64	0.49	0.629
Digit Span	0.005	0.01	0.59	0.555
GJT(ungrammatical)	-0.001	0.12	-0.01	0.994

Note. GJT = grammaticality judgment task.
$p < .05$

results suggest that with increasing LLAMA F scores, the performance on correct items may even get disadvantaged. On the one hand, this unexpected observation might be attributed to analytically-minded learners' tendency to overthink GJTs (i.e., discerning errors where there are none). However, it needs to be stressed that the grammatical items reached ceiling scores in the study, rendering the variation in the data very limited. Also, the aggregated scores in all correct items were not significantly related to LLAMA F scores. Consequently, the results related to grammatical items need to be approached with caution. Moreover, largely in line with previous research, this study did not show links between explicit learning aptitude and online processing of L2 grammar, measured by the SPR task. Similarly, no such relationship was reported by Suzuki and DeKeyser (2017). In sum, the findings related to explicit learning abilities largely corroborated the results of previous research (Granena & Yilmaz, 2018), which showed links between explicit cognitive abilities and L2 knowledge and learning. Additionally, the results showed that this effect might hold even when general L2 proficiency is controlled for.

The second area explored in the study was the role of working memory in L2 processing. Even though WM is a factor posited to be involved in acquiring L2 (Juffs & Harrington, 2011; Linck et al., 2014), in the present study, it did not predict the results of either the GJT or SPR task. In the case of offline language processing, none of the analyses, varying in an array of structures considered, revealed any links between WM and scores in the GJT. Also, reading times, assumed to reflect implicit processing, were not related to WM capacity. As mentioned above, Faretta-Stutenberg and Morgan-Short (2018) found that the role of WM in L2 processing was limited to changes in online processing, when L2 was acquired through immersion. Their study did not show any relationship between WM and the GJT in either study-abroad (i.e., immersion) or at-home (i.e., classroom) settings. In the present study, as follows from the demographic background questionnaire, the majority of the participants learned English in the classroom context. Moreover, explicit form-focus instruction is generally favoured in mainstream L2 instruction in Poland (Pawlak, 2006). Thus, the present results might provide further evidence that WM is not as important in the classroom settings, which place lower cognitive demands on real-time L2 processing, as it is in immersion settings (see also Tokowicz et al., 2004). Moreover, as suggested by Suzuki and DeKeyser (2017) WM might not be particularly relevant for mastering specific grammatical structures, as opposed to developing general L2 proficiency (but see Erçetin & Aptekin, 2013, for a contrasting finding). To further explore links between WM and development of L2 grammar, future research should utilize other WM tasks, such as the reading span task, which might be more closely related to knowledge of L2 grammar than a digit span task (see Juffs & Harrington, 2011).

Finally, as regards online processing of relative pronouns, none of the predictor variables accounted for participants' performance on the SPR task. Quite surprisingly, even general L2 proficiency did not predict the reading times in the online task. Moreover, in contrast to previous research (Pélissier et al., 2017; Suzuki & DeKeyser, 2017) no relationship was found between the offline (explicit) knowledge of the target structure and its online processing. The discrepancy between the findings by Suzuki and DeKeyser (2017) and the present study might be explained on the grounds that the latter study considered composite scores in three different structures, whose processing was tapped into by three psycholinguistic measures of implicit knowledge (an SPR task was one of them), while in the current study, a single target structure was presented via a single online task. Moreover, Suzuki and DeKeyser (2017) focused on the construct of automatized explicit knowledge measured by time-pressured tasks, while the current study used an untimed task, for which speedy processing was not essential. Hypothetically, the more automatized explicit knowledge is, the more relevant it becomes for the formation of implicit knowledge (see DeKeyser, 2017). Finally, in line with much of the previous research, in the present study, no links were found between online processing and explicit learning aptitude or WM capacity. Possibly, other cognitive skills, such as statistical learning abilities might determine implicit processing of L2 grammar (e.g., Granena, 2013).

Limitations and Future Directions

The current study was not free from limitations. First, as discussed above, in the offline GJT, the target contrast between relative pronouns turned out to be rather easy for the participants, producing ceiling effects in the case grammatical items. Second, the design lacked a measure of implicit learning aptitude (e.g., a serial reaction time task) to better model the network of possible interactions between aptitudes and modes of L2 processing. Third, due to the correlational nature of the study, it was not possible to infer cause-and-effect relationships from the data. Furthermore, the design did not allow for controlling for many exposure-related variables, such as the quantity of input or quality of instruction. However, arguably, substantial variation related to L2 exposure was controlled for by the implementation of English proficiency scores into the statistical models.

Future research should focus on more diversified L2 structures varying in grammatical complexity as well as degrees of similarity to L1 (to explore the role of cross-linguistic influences). Moreover, the designs should incorporate other measures of cognitive aptitudes, including other complex span WM tasks (e.g., a reading span task) and tasks tapping into implicit learning aptitude (e.g., the SRT task). Finally, findings of correlational research should be gauged taking into account the results of corresponding experimental studies. Whereas the former designs allow for investigating L2 knowledge developed in naturalistic conditions, the latter impose strict control over many exposure-related variables.

Nevertheless, despite the limitations, the current study confirmed that explicit learning aptitude plays an important role in developing offline but not online knolwedge of L2 grammar. While previous research focused on learners coming from specific educational programmes or immersion learners; in this study, the correlation between explicit learning aptitude and explicit knolwedge of L2 was replicated on a considerably diversified sample of L2 learners. Additionally, no relationships were found between knowledge of grammar and WM in both online and offline tasks.

FOOTNOTES

[1] Certain ERP components are considered measures of online language processing, however, the extent to which they reflect implicit language knowledge is a matter of debate (Morgan-Short et al., 2015).
[2] The reliability was computed for segment $n + 1$, which showed the most robust processing effects. Due to differences in the structure of stimuli, the analysis was carried out separately for "which" and "what" items.

ACKNOWLEDGEMENTS

This research was partially supported by the internal grant No. DSM11930006/2018 from the Faculty of Psychology, University of Warsaw, and partially by grant No. 2016/21/B/HS6/01129 from the National Science Centre Poland (PI: Agnieszka Otwinowska-Kasztelanic, University of Warsaw). The preparation of the manuscript was supported by the University of Economics and Human Sciences in Warsaw..

I would like to thank Agnieszka Otwinowska-Kasztelanic, Agata Ambroziak and Katarzyna Kutyłowska for their help in managing data collection for this study. Many thanks to Jakub Szewczyk for consulting the statistical analyses.

REFERENCES

Abrahamsson, N., & Hyltenstam, K. (2008). The robustness of aptitude effects in near-native second language acquisition. *Studies in Second Language Acquisition, 30*, 481–509. doi: 10.1017/S027226310808073X

Baddeley, A. (2003). Working memory and language: An overview. *Journal of Communication Disorders, 36*, 189–208. doi: 10.1016/S0021-9924(03)00019-4

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language, 68*, 255–278. 10.1016/j.jml.2012.11.001. doi: 10.1016/j.jml.2012.11.001

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

Brzeziński, J., Gaul, M., Hornowska, E., Jaworowska, A., Machowski, A., & Zakrzewska, M. (2004). *WAIS-R (PL)—Skala Inteligencji Wechslera dla Dorosłych—Wersja Zrewidowana*. Pracownia Testów Psychologicznych.

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

De Graaff, R. (1997). The Experanto experiment: Effects of explicit instruction on second language acquisition. *Studies in Second Language Acquisition, 19*, 249–276. doi: 10.1017/S0272263197002064

DeKeyser, R. (2000). The robustness of critical period effects in second language acquisition. *Studies in Second Language Acquisition, 22*, 499–533. doi: 10.1017/S0272263100004022

DeKeyser, R. (2012). Interactions between individual differences, treatments, and structures in SLA. *Language Learning, 62*, 189–200. doi: 10.1111/j.1467-9922.2012.00712.x

DeKeyser, R. (2017). Knowledge and skill in SLA. In S. Loewen & M. Sato (Eds.), *Handbook of instructed second language acquisition* (pp. 15–32). Routledge.

Doughty, C. J. (2019). Cognitive language aptitude. *Language Learning, 69*, 101–126. doi: 10.1111/lang.12322

Erçetīn, G., & Alptekīn, C. (2013). The explicit/implicit knowledge distinction and working memory: Implications for second-language reading comprehension. *Applied Psycholinguistics, 34*, 727–753. doi: 10.1017/S0142716411000932

Faretta-Stutenberg, M., & Morgan-Short, K. (2018). The interplay of individual differences and context of learning in behavioral and neurocognitive second language development. *Second Language Research, 34*, 67-101. doi: 10.1177/0267658316684903

Foryś-Nogala, M. (2019). *The role of implicit and explicit training in shaping declarative and procedural representations of foreign language syntax* [Unpublished doctoral thesis. Faculty of Psychology, University of Warsaw, Poland].
https://depotuw.ceon.pl/handle/item/3381

Gozdawa-Gołębiowski, R., & Opacki, M. (2018). Recurrent strings in corpus-based pedagogical research: A reappraisal of the field. *Glottodidactica. An International Journal of Applied Linguistics, 45*, 134–149. doi: 10.14746/gl.2018.45.2.07

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

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*, 311–343. doi: 10.1177/0267658312461497

Granena, G., & Yilmaz, Y. (2018). Aptitude-treatment interaction in l2 learning: A research synthesis. *Studies in English Education, 23*, 803–830.

Hulstijn, J. (2002). Towards a unified account of the representation, processing and acquisition of second language knowledge. *Second Language Research, 18*, 193–223. doi: 10.1191/0267658302sr207oa

Jiang, N., Novokshanova, E., Masuda, K., & Wang, X. (2011).

Morphological congruency and the acquisition of L2 morphemes. *Language Learning, 61*, 940–967. doi: 10.1111/j.1467-9922.2010.00627.x

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

Just, M. A., Carpenter, P. A., & Woolley, J. D. (1982). Paradigms and processes in reading comprehension. *Journal of Experimental Psychology. General, 111*, 228–238.

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

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software, 82*, 1–26. doi: 10.18637/jss.v082.i13

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*, 530–566. doi: 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*, 861–883. doi: 10.3758/s13423-013-0565-2

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

Morgan-Short, K., Faretta-Stutenberg, M., & Bartlett-Hsu, L. (2015). Contributions of ERP research to issues in implicit/explicit learning. In P. Rebuschat (Ed.), *Implicit and explicit learning of languages* (pp. 349–383). John Benjamins Publishing Company.

Paradis, M. (2009). *Declarative and procedural determinants of second languages.* John Benjamins Publishing Company.

Pawlak, M. (2006). *The place of form-focused instruction in the foreign language classroom*. Adam Mickiewicz University Press.

Pawlak, M., & Biedroń, A. (2021). Working memory as a factor mediating explicit and implicit knowledge of English grammar. *Annual Review of Applied Linguistics, 41*, 118–125. doi: 10.1017/S0267190521000052

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: 10.3758/s13428-018-01193-y

Pélissier, M., Krzonowski, J., & Ferragne, E. (2017). Does better explicit knowledge of a morphosyntactic structure guarantee more native-like electrophysiological processing? An ERP study with French learners of English. *SHS Web of Conferences, 38*, 00002. doi: 10.1051/shsconf/20173800002

Powell, M. J. (2009). The BOBYQA algorithm for bound constrained optimization without derivatives. *Cambridge NA Report NA2009/06, University of Cambridge.*

R Core Team. (2020). R: A language and environment for statistical computing. *Foundation for Statistical Computing*. https://www.R-project.org/

Skehan, P. (1998). *A cognitive approach to language learning*. Oxford University Press.

Suzuki, Y. (2017). Validity of new measures of implicit knowledge: Distinguishing implicit knowledge from automatized explicit knowledge. *Applied Psycholinguistics, 38*, 1229–1261. doi: 10.1017/S014271641700011X

Suzuki, Y., & DeKeyser, R. (2015). Comparing elicited imitation and word monitoring as measures of implicit knowledge. *Language Learning, 65*, 860–895. doi: 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*, 747–790. doi: 10.1111/lang.12241

Tokowicz, N., Michael, E. B., & Kroll, J. F. (2004). The roles of study-abroad experience and working-memory capacity in the types of errors made during translation. *Bilingualism: Language and Cognition, 7*, 255–272. doi: 10.1017/S1366728904001634

Trueswell, J. C., Tanenhaus, M. K., & Garnsey, S. M. (1994). Semantic influences on parsing: use of thematic role information in syntactic ambiguity resolution. *Journal of Memory and Language, 33*, 285–318. doi: 10.1006/jmla.1994.1014

Vafaee, P., Suzuki, Y., & Kachisnke, I. (2017). Validating grammaticality judgment tests: Evidence from two new psycholinguistic measures. *Studies in Second Language Acquisition, 39*, 59–95. doi: 10.1017/S0272263115000455

Wen, Z. (2016). Working memory and second language learning: Towards an integrated approach. *Multilingual Matters.*

Wen, Z., Biedroń, A., & Skehan, P. (2017). Foreign language aptitude theory: Yesterday, today and tomorrow. *Language Teaching, 50*, 1–31. doi: 10.1017/S0261444816000276

Williams, J. N. (2009). Implicit learning in second language acquisition. In W. C. Ritchie & T. K. Bhatia (Eds.), *The new handbook of second language acquisition* (pp. 319–352). Brill.

Yilmaz, Y. (2013). Relative effects of explicit and implicit feedback: The role of working memory capacity and language analytic ability. *Applied Linguistics, 34*, 344–368. doi: 10.1093/applin/ams044

Yilmaz, Y., & Granena, G. (2021). Implicitness and explicitness in cognitive abilities and corrective feedback: A double dissociation? *Studies in Second Language Acquisition, 43*, 523–550. doi: 10.1017/S0272263120000601

RECEIVED 09.11.2019 | ACCEPTED 22.12.2021