Studies in Second Language Acquisition, 2016, **38**, 239–263.
doi:10.1017/S0272263115000509

LANGUAGE APTITUDE AND GRAMMATICAL DIFFICULTY

An EFL Classroom-Based Study

Şebnem Yalçın
Boğaziçi University

Nina Spada
University of Toronto

This study investigates the relationship between foreign language aptitude and the learning of two English structures defined as easy or difficult to learn. Using a quasiexperimental design, 66 secondary-level learners of English as a foreign language from three intact classes were provided with four hours of instruction on the *passive* (a difficult structure) and the *past progressive* (an easy structure). Language aptitude was measured using the LLAMA Aptitude Test (Meara, 2005). Language outcomes were measured with a written grammaticality judgment and an oral production task. The results revealed that one of the aptitude components, grammatical inferencing, contributed to learners' gains on the *passive* but not the *past progressive* on the written measure. Another component of aptitude, associative memory, contributed to learners' gains on the *past progressive* on the oral measure. The results provide support for the claim that different components of aptitude contribute to the learning of difficult and easy L2 structures in different ways. There is also support for the proposal that different components of aptitude may be involved at different stages of language acquisition (Skehan, 2002).

Correspondence concerning this manuscript should be addressed to Şebnem Yalçın, Boğaziçi University Faculty of Education, Department of Foreign Language Education, Bebek 34342 İstanbul, Turkey. E-mail: sebnem.yalcin@boun.edu.tr

© Cambridge University Press 2016

https://doi.org/10.1017/S0272263115000509 Published online by Cambridge University Press

INTRODUCTION

In the second language acquisition (SLA) literature, language aptitude has been documented as one of the most consistent predictors of second language (L2) success (Dörnyei & Skehan, 2003; Ehrman & Oxford, 1995), both in classroom contexts (Harley & Hart, 1997, 2002; Kormos & Sáfár, 2008; Wesche, 1981) and in natural contexts (Reves, 1983 cited in Dörnyei & Skehan, 2003). Aptitude has also been found to play a role across different L2 instructional conditions (e.g., explicit and implicit) (Erlam, 2005; Robinson, 2002; Sheen 2007a, 2007b), different age groups (DeKeyser, 2000; Harley & Hart, 1997, 2002), and in relationship to the learning of different language features (Robinson, 2002). The focus of this research is on the latter—to investigate the extent to which aptitude predicts more (or less) successful learning depending on whether the target feature is easy or difficult to learn.

Conceptualization and Measurement of Aptitude

Language aptitude has been defined as a "specific talent for learning foreign languages that exhibits considerable variation between learners" (Dörnyei & Skehan, 2003, p. 613). It is considered to be a multicomponential construct consisting of a combination of different cognitive abilities. Originally, Carroll (1981) proposed four subcomponents of language aptitude: phonetic coding ability (the ability to identify distinct sounds, to form associations between sounds and the symbols representing them), grammatical sensitivity (the ability to recognize the grammatical functions of words in sentences), inductive language learning ability (the capacity to infer or induce the rules governing a set of language materials and extrapolate to new language), and associative memory (the ability to learn associations between sounds and meanings rapidly and efficiently). These components have been investigated ever since in studies exploring the role of aptitude in L2 learning (e.g., Abrahamsson & Hyltenstam, 2008; de Graaff, 1997; Ehrman & Oxford, 1995; Harley & Hart, 1997, 2002; Robinson, 2002).

Over the years, other conceptualizations of language aptitude have been proposed. For example, Skehan (1986, 2002) suggested three as opposed to four components arguing that Carroll's (1981) original categories of grammatical sensitivity and inductive language learning ability could be combined into one. Skehan proposed a model consisting of auditory processing, language analysis, and memory components. More recently in a factor analytic study of the contributions of L1 cognitive ability, L2 aptitude, and L2 affective measures to predict oral and written

https://doi.org/10.1017/S0272263115000509 Published online by Cambridge University Press

L2 proficiency, Sparks, Patton, Ganschow, & Humbach (2011) observed that four factors explained 76% of the variance in learners' scores: (a) language analysis composed of L1 and L2 comprehension, grammar, vocabulary, and inductive language learning; (b) phonology/orthography composed of L1 and L2 phonemic coding and phonological processing; (c) memory consisting of L1 intelligence and L2 paired-associate learning measures, and (d) self-perceptions of language skills, composed of L2 motivation and L2 anxiety measures. Their findings thus support the componential view of L2 aptitude and confirm the long-term relationship that exists between L1 and L2 learning. Their study also pointed to a "puzzling result . . . that the measure of rote memory—the Modern Language Aptitude Test (MLAT) Paired Associates subtest—loaded onto the same factor as the IQ test" (Sparks et al., 2011, p. 267). This, combined with the observation that the MLAT Paired Associates subtest and rote memory components of aptitude have been found to be the least effective predictors of L2 success in other studies (Carroll, 1990; Rysiewicz, 2008), has led some researchers to argue that working memory (WM) rather than rote memory may be a more central component of aptitude (Harrington & Sawyer, 1992; Miyake & Friedman, 1998).[1]

One of the issues regarding the conceptualization of aptitude is whether it is a relatively stable trait or changes over time. Researchers who support a dynamic conceptualization of aptitude (e.g., Dörnyei, 2005; Robinson, 2001, 2012; Skehan, 2002, 2010) propose that learners possess different aptitude profiles and that they may access different groups of cognitive abilities at different ages. Some supporting evidence comes from research showing that the memory component of aptitude is more related to success with younger L2 learners whereas the language analytic component is a stronger predictor of successful L2 learning with older learners (Harley & Hart, 1997, 2002).

It has also been suggested that aptitude abilities vary in relation to learner's development. For example, Skehan's (2002) aptitude profile model suggests nine developmental stages with four main levels: *noticing* (Stage 1), *patterning* (Stages 2–5), *controlling* (Stages 6–8), and *lexicalising* (Stage 9). *Noticing* refers to the first stage in the acquisition of a language feature in which learners encounter, or are introduced to, a new structure and their attention is directed to specific aspects of the language system (Schmidt, 1995). *Patterning* represents the stage during which learners are expected to make generalizations based on the language input. *Controlling* involves learners efforts to proceduralize knowledge for fluent performance. And *lexicalizing* represents the final stage during which learners move beyond rule-based processes and are capable of accessing and using language quickly and without error. Skehan argues that at this stage, learners need to access a repertoire of linguistic expressions quickly to perform at nativelike levels emphasizing

https://doi.org/10.1017/S0272263115000509 Published online by Cambridge University Press

the important role for memory. According to the model, different components of aptitude are relevant at different stages. For example, grammatical sensitivity and inductive language learning are thought to come into play at the beginning of the second stage (*pattern identification*), whereas phonetic sensitivity and WM are relevant at the lexicalizing stage (see Skehan, 2002, for more detailed discussion of this model).

In terms of measurement, the use of the MLAT (Carroll & Sapon, 1959) to assess L2 learners' aptitude has dominated the field for more than five decades. This commercially available test consists of five subtests intended to measure the four aptitude components originally proposed by Carroll (i.e., phonemic coding ability, grammatical sensitivity, rote learning ability, inductive learning ability). A commercially available alternative to the MLAT is the Pimsleur Language Aptitude Battery (PLAB: Pimsleur, 1966) that was developed for high school children. It is similar to the MLAT with respect to the cognitive abilities measured but also includes a questionnaire to measure learners' motivation levels. Other aptitude tests include the Cognitive Ability for Novelty in Acquisition of Language (CANAL-F) developed by Grigorenko, Sternberg, and Ehrman (2000), which is based on a cognitive theory of knowledge acquisition that stresses the importance of coping with novelty in learning (Sternberg, 1999). An aptitude measure currently in development and targeted to advanced proficiency language learners is the Hi-LAB (Linck et al., 2013). This test emphasizes the role of memory (i.e., phonological short-term memory, associative memory, and long-term memory), implicit learning, and auditory discrimination. Another relatively recent measure, the LLAMA aptitude test, developed by Meara (2005) and modeled on the MLAT, was used in the present study and is described in more detail in the following text. LLAMA was used in this study because it is language independent and is easily administered by computer. Its use in a pilot study also indicated that the user-friendly characteristics including computer graphics made it motivating for the adolescent target population.

Language Aptitude, Type of Instruction, and L2 Learning

It has been acknowledged that research investigating the mediating role of aptitude in the effectiveness of L2 instruction can provide valuable information for L2 pedagogy (DeKeyser, 2012; Robinson, 2001, 2012). Two models of language aptitude formulated by Skehan (2002) and Robinson (2001) have paved the way for exploring the possibility of playing to learners' strengths by tailoring L2 instruction to learners' needs and preferences that will hopefully boost L2 learning. This line of research is referred to as aptitude-treatment-interactions (ATI). One of the earliest

https://doi.org/10.1017/S0272263115000509 Published online by Cambridge University Press

classroom studies to investigate aptitude in relation to type of L2 instruction was conducted by Wesche (1981). She reported that adult learners matched with three different types of L2 instructional treatment (i.e., audio-visual, functional, and analytical) that were deemed to be compatible with their aptitude profiles and learning styles did better than learners who were not matched with a compatible type of instruction.

In a more recent study, Erlam (2005) investigated the relationship between aptitude and three different instructional interventions (i.e., deductive, inductive, and structured input) with French language learners who were taught direct object pronouns in French. The results indicated that deductive instruction minimized the effects of aptitude on learning outcomes with no relationship found between the L2 gain scores and language aptitude scores. In the structured input group, however, learners with greater language analytical ability (LAA) and WM capacity benefited more from instruction. Finally, in the inductive group only analytical ability made a difference in terms of learners' performance. In a study with learners of Spanish, Hwu and Sun (2012) examined how language aptitude interacted with two L2 instructional conditions in the learning of the verb *gustar* (to like; see also Leow et al., this issue). One group of university-level students received explicit deductive L2 instruction while the other received explicit inductive instruction. The results showed that learners in both instructional groups significantly improved and high aptitude learners benefitted more from both types of instruction than low aptitude learners.

In two recent studies VanPatten and colleagues questioned the central role assigned to aptitude especially in adult L2 learning. VanPatten and Smith (2015) investigated the predictive power of aptitude in L2 instruction without rule learning (i.e., implicit instruction). The results showed that aptitude, as measured by grammatical sensitivity, was not a factor in unconscious L2 reading times for grammatical and ungrammatical sentences (their study focused on head directionality as well as case marking in Japanese L2). Therefore, no role for language aptitude was reported in the study. The researchers argued that when L2 learning is not defined as the learning of explicit rules but instead as the internalization of a generative system, then aptitude may not be a significant factor in acquisition. In another study VanPatten, Collopy, Price, Borst, and Qualin (2013) used processing instruction and targeted four languages (i.e., Spanish, German, Russian, and French). The researchers focused on the correct processing of sentences rather than any kind of rule learning. The findings revealed that except for one group (i.e., German, explicit instruction), there was no significant correlation between language aptitude and trials-to-criterion (i.e., how long it took participants to begin processing correctly). They state that when the aim is not the internalization of a rule, grammatical sensitivity is probably not relevant to language learning.

https://doi.org/10.1017/S0272263115000509 Published online by Cambridge University Press

In research that has explored potential aptitude effects in relation to different types of corrective feedback (CF), Sheen (2007a, 2007b) reported significant correlations between students' LAA and metalinguistic correction, but not with oral recasts (i.e., implicit reformulations of learners' errors). In a laboratory study, Yilmaz (2012) explored the role of WM and LAA in mediating how learners benefit from different types of feedback (i.e., explicit correction, recast, and no feedback). The results indicated that explicit correction ruled out individual differences in WM and LAA. That is, both high and low analytical ability learners benefitted from explicit correction. Recasts were observed not to be effective for one of the language structures targeted (locatives), but they were found to be effective for another structure (i.e., plurals) but only among the high LAA learners. These results are consistent with Sheen's (2007a, 2007b) finding that explicit correction is related to LAA.

In other aptitude and CF interaction research, different findings have been reported. Trofimovich, Ammar, and Gatbonton (2007) found that LAA determined the effectiveness of computerized recasts on L2 morphosyntactic accuracy. Similarly, in Li (2013) the benefits of implicit correction on classifiers in Chinese were found to be sensitive to differences in LAA. In interpreting these findings, Li (2013) hypothesized that "other things being equal, language analytic ability is implicated in implicit conditions in the learning of easy, transparent structures that are within one's processing capacity, and in explicit conditions in the learning of hard, opaque structures where the internalization of available metalinguistic information sets heavy processing demands on internal cognitive resources" (p. 648). Li argued that aptitude is called upon and is useful when the instruction is implicit and the target structure is easy, and when the instruction is explicit and the L2 structure is difficult. Li (2015b) provided some evidence to support this hypothesis in a study with Chinese learners who received recasts and metalinguistic feedback on the Chinese perfective marker (i.e., *–le*)—a form with an opaque form-meaning mapping and a complex rule explanation. The results indicated that LAA was not sensitive to the effects of recasts, but that it significantly predicted the effects of metalinguistic feedback. It seems that for the difficult structures to benefit from CF, high aptitude and explicit feedback are prerequisites. Skehan (2015) also suggested that the features of the target structure (i.e., redundancy and saliency) likely play an important role when it comes to learners with different aptitude profiles benefitting from CF stating that "higher aptitude seems to help a focus-on-form- to be 'dug out'" (p. 14). That is, learners with higher aptitude can rely on their skills to notice and figure out regularities in the input when it is not structured for them. This led Skehan to speculate that aptitude may be more important for early stages of L2 development (in his set of acquisitional stages) where noticing and pattern identification occur.

https://doi.org/10.1017/S0272263115000509 Published online by Cambridge University Press

Grammatical Difficulty and Language Aptitude

According to DeKeyser (2012), investigations of aptitude and L2 structure interaction have the potential to demystify the processes through which different structures are learned. Yet, little research has investigated the role that aptitude might play when L2 learners are faced with the task of learning language features that differ in terms of their difficulty or complexity. L2 difficulty has been investigated in instructed SLA research based on the assumption that the effectiveness of a particular type of L2 instruction (e.g., explicit or implicit) may depend on the language feature targeted for instruction (DeKeyser, 1998; Doughty & Williams, 1998; Housen, Van Daele, & Pierrard, 2005; Spada & Tomita, 2010). In order to address aptitude by L2 structure interaction in the present study the term *L2 difficulty* is used as a broader concept than complexity. Whereas complexity is more related to the inherent properties of the target features, difficulty refers to the "relative ease or difficulty (in terms of cognitive effort and resources deployed) with which language features are processed in L2 learning or use" (Housen, 2014, p. 2). There are subjective and objective determinants of difficulty. Subjective determinants refer to individual learner factors such as aptitude, motivation, L2 proficiency, and cognitive maturity. Thus, a language structure that may be moderately difficult for one learner may be easy for another learner with more language learning aptitude or language learning experience. This led DeKeyser (2005) to define rule difficulty as "the ratio of the rule's inherent linguistic complexity to the student's ability to handle such a rule" (p. 331). Objective determinants of difficulty are related to the ease (or lack of ease) regarding learners' access to a language feature in the input. This includes, for example, its perceptual saliency (e.g., *–ing* in English more salient than third person *–s*) and frequency (e.g., plural *–s* in English more frequent in the input than possessive *–s*) (Goldschneider & DeKeyser, 2001).[2]

Little research has investigated the role that aptitude might play when L2 learners are faced with the task of learning language features that differ in terms of their difficulty. In a study by Robinson (2002), three L2 structures in Samoan (i.e., locative, ergative, and incorporation) were investigated in relation to three types of instruction (i.e., implicit, explicit, and incidental). The three structures varied in terms of difficulty with the ergative rule considered to be the most difficult due to its novel and abstract nature. The language gains were measured through computerized written and oral grammaticality judgment tasks (GJTs) and guided sentence production tasks. WM was measured using Osaka and Osaka's (1992) reading span task and aptitude was measured using Sasaki's (1996) Language Aptitude Battery for the Japanese. The results revealed that aptitude and WM interacted with rule difficulty in

https://doi.org/10.1017/S0272263115000509 Published online by Cambridge University Press

the incidental condition. That is, performance on one of the target rules (for locative sentences) was positively related to aptitude but not to WM whereas performance on another target rule (for ergative structures) was negatively related to aptitude. The third rule (i.e., incorporation) was not related to aptitude or WM. Robinson concluded that "adult incidental L2 learning is invariably affected by the complexity, and the nature of the rule to be learned, and individual differences in cognitive abilities interact with rule complexity and rule type during incidental learning" (2002, p. 260).

To our knowledge no other research has investigated the role that the target language feature plays in studies exploring interactions between aptitude, type of instruction, and L2 learning. While this study did not manipulate type of instruction, it intends to contribute to research in this domain by investigating the following question in a classroom setting: What role does aptitude play in the learning of two language features that differ in terms of their difficulty? This question leads to the following hypothesis, based on the findings of the research discussed in the preceding text: Aptitude will play a greater role in the explicit learning of a difficult L2 structure and a lesser role in the explicit learning of an easy structure.

METHODS

Context and Participants

The participants were 66 EFL learners enrolled in the eighth grade (13 to 14 years old) at a private[3] secondary school in Turkey. The learners were at a preintermediate level of English proficiency as determined by their teachers' expert ratings. For all participants, English was an obligatory subject of which they received seven hours of instruction per week. Although the general aim of English language teaching at this school was to improve L2 communicative language skills, there was also a focus on formal instruction due to a nationwide exam, which measures explicit knowledge of grammar.

Target Structures

The study examines two target structures that are hypothesized to differ in terms of difficulty: the passive voice and past progressive tense in English. One of the ways in which difficulty was operationalized was with respect to structural complexity, following Hulstijn and de Graaff (1994). That is, the passive is considered to be the "difficult" feature

https://doi.org/10.1017/S0272263115000509 Published online by Cambridge University Press

because its production involves a higher number of grammatical operations compared to the past progressive. Its lack of frequency in the input (Hinkel, 2004) is another reason contributing to the difficulty of the passive. The status of the passive as a difficult structure is also supported by its relatively late acquisition for L1 learners of English (Kirby, 2010).

The past progressive was characterized as the "easy" structure because of its transparent form-meaning relationship and frequency in the input (Collins, Trofimovich, White, Cardoso, & Horst, 2009; Révész, 2009). Collins et al. (2009) also noted that the past progressive is more available and accessible in the input as it occurs across a wide range of common and highly frequent verbs. In addition, according to Révész (2009) the past progressive is physically salient as it is realized by a free morpheme (was/were) and syllabic bound morpheme (*–ing*). All progressive markers furthermore appear as intact syllables, which makes the past progressive also more phonetically salient. There are furthermore no allomorphs for the *–ing* marker that would reduce its transparency. In contrast, the variety of allomorphs in the formation of the passive may render it more difficult to learn than the progressive. In addition to the "objective criteria" of formal complexity, input frequency, and salience differences, we also considered the "perceived difficulty" of the two features. We elicited the participants' opinions about the difficulty of six L2 structures (i.e., tenses and passive voice) that were taught in their regular English curriculum. They were asked to rate the difficulty of each of the target structures on a five-point scale. The learners' responses indicated that the passive voice was considered to be a more difficult structure to learn when compared with the past progressive and a paired-samples t-test indicated that this difference was statistically significant t (59) = -7.45, $p < .001$).

INSTRUCTIONAL TREATMENT

One of the researchers replaced the EFL teacher to teach the target structures. To establish familiarity with the students, the researcher spent several weeks at the school beforehand and thus the students came to know her well and were comfortable with her before the intervention began. The two target structures were taught for four hours each and testing took approximately two additional hours. Learners followed their regular school curriculum during the experiment. They were about to study the passive voice when the intervention started toward the end of the first semester. As it happened, the students had already been introduced to the past progressive at the beginning of the semester. After obtaining informed consent, the learners were pretested with the language measures to control for previous learning of the two

https://doi.org/10.1017/S0272263115000509 Published online by Cambridge University Press

target structures in the study (i.e., GJT and oral production tasks [OPT]) before the instruction began. Consistent with Spada's (1997) definition of form-focussed instruction, the instructional materials provided learners with explicit information about the formal properties of the target features and they were given opportunities to practice them during communicative activities. CF was provided both implicitly and explicitly throughout all activities. (See online Supporting Information for sample instructional materials.)

Language and Aptitude Measures

Language Aptitude Test. LLAMA (Meara, 2005) is a computerized aptitude test that includes four subcomponents: (a) vocabulary learning task: LLAMA B, (b) sound recognition task: LLAMA D, (c) sound-symbol correspondence task: LLAMA E, and (d) grammatical inferencing task: LLAMA F. LLAMA B measures the ability to learn relatively large amounts of vocabulary in a short time and is therefore considered to be a measure of rote (associative) memory. The words are adapted from a Central American language. LLAMA D is a sound recognition test designed to test if participants can identify short pieces of previously heard sound sequences in new sequences. The sound sequences are based on the names of objects in a British-Columbian Indian language. LLAMA E presents a set of 24 recorded syllables with a transliteration of these syllables in an unfamiliar alphabet. The learners' task is to work out the relationship between the sounds heard and the writing system. Learners hear the target syllables through headphones. LLAMA F is considered to be similar to the *words in sentences* component in the MLAT. It presents learners with sentences in an unknown language and requires them to figure out the morphosyntactic rules. While the LLAMA aptitude test has been used quite extensively in SLA research (Granena & Long, 2013; Yalçın, Çeçen, & Erçetin, forthcoming; Yilmaz, 2012), it has not yet undergone systematic validation. One exception is Granena (2013) who carried out an exploratory validation of the test. The results indicated that LLAMA E, F, and B appear to be measuring aptitude for explicit language learning because according to Granena, they all include explicit cognitive and memory processes (e.g., study phase, problem-solving strategies, rehearsal, rote learning, explicit associative learning). The results for LLAMA D, however, suggest that it measures aptitude for implicit language learning because it does not call for the use of analytical abilities. Learners are simply exposed to sound sequences but not asked to work out rules or relationships nor given the time or opportunity to do so.

https://doi.org/10.1017/S0272263115000509 Published online by Cambridge University Press

Grammaticality Judgment Tasks. To measure explicit L2 knowledge, two written and untimed GJT were developed for the two target structures. Following Ellis (2009), the GJT was hypothesized to be a measure of explicit knowledge because it focuses on form, is not performed under time pressure, and encourages a high degree of awareness. There were 40 items in the past progressive GJT. Sixteen of them represented the prototypical context (i.e., indicating that an event or action was in progress at a particular reference time in the past; Swan, 1995) and 24 represented the context where the past progressive is used to describe a longer "background" action. The 16 items representing the prototypical use of past progressive (e.g., *She was painting the house at 11 last night*) were adapted from Révész (2007). In order to determine whether learners had control over all contexts for this tense-aspect form the additional 24 items included *while* and *when* time phrases (e.g., *My father was still sleeping when I left for school* or *While I was going to the yoga class I stopped to buy a magazine*). Twenty-one items were grammatical and 19 were ungrammatical.[4] (See online Supporting Information for sample GJT items.) Learners were asked to decide whether the sentence was grammatical and, if so, provide the correct form. Two versions of the GJT were used, differing only in the order of items. The participants had three response options: *Correct, Incorrect*, and *I do not know*. Following Spada, Jessop, Suzuki, Tomita, and Valeo (2014), partial scoring was used to take into consideration interlanguage development. Participants received full points (i.e., 3 points) for identifying a grammatical and ungrammatical past progressive sentence and for correcting it in the case of ungrammaticality. They received two points if they provided the present progressive form instead of the past (e.g., "While they are eating in a café the waiter dropped the dishes"), they received one point if they provided a bare progressive with no auxiliary (e.g., "She eating a sandwich at 9 o'clock last night"). They received zero if they changed the sentence to a nonprogressive form (see Yalcin, 2013 for more details on the scoring procedures).

The GJT for the passive also included 40 items. Sixteen items were grammatical and 24 were ungrammatical. The items were adapted from Spada et al. (2014) and included passive statements formed with the present tense (16 items), past tense (16 items), and the present perfect tense (eight items). Both GJTs were paper and pencil tests administered without any time constraints. Partial scoring was also used for the passive GJT. When learners accurately corrected an ungrammatical passive sentence (i.e., a sentence missing either the auxiliary 'be', or the past participle '-ed'), they received two points. If they failed to provide the past participle form of the main verb, while correcting an ungrammatical sentence, they were given one point (e.g., "The tires on the car were change"). The provision of "be" auxiliary and the past participle

https://doi.org/10.1017/S0272263115000509 Published online by Cambridge University Press

form of the main verb were required components to obtain full points. Minor spelling errors for past participle forms were ignored as long as there was an attempt to provide the form (see Yalcin, 2013 for more details on the scoring procedures).

Oral Production Tasks. To measure learners' productive use of the L2 structures, two picture-cued OPTs were developed, one for each structure. Both OPTs were administered individually and involved a research assistant interacting with a learner using a Microsoft PowerPoint presentation of the pictures for each task. The past progressive OPT was designed for this study and contextualized as a hypothetical scenario that placed the test-taker in the role of someone who could confirm the alibi of a person convicted of a crime. In the passive OPT learners were asked to tell a story about a package that was lost in the mail (Spada et al., 2014). Both OPTs were piloted with a similar population of learners prior to the study and revised accordingly. The OPT is considered to be a task in which learners draw more on their implicit L2 knowledge[5] given that it was designed to direct the learners' attention to meaning not form in a one-on-one communicative exchange. (See online Supporting Information for the OPTs.) Scoring of the OPTs was based on accurate use of target structures in the obligatory contexts. Learners received full points for accurate use of target structures in obligatory context, no points for incorrect form, and partial points for interlanguage forms. For example, on the passive OPT, if the learner produced a correct passive sentence, three points were given. If the learner produced a passive sentence with errors in agreement and/or past participle (e.g., "The package was sended to Canada."), two points were given. If an active instead of a passive sentence was produced (e.g., "They put the package on the truck."), one point was given (see Yalcin, 2013 for more details on the scoring procedures).

This quasiexperimental pre-/posttest instructional intervention study took place over six weeks with three intact classes. Data collection took place during the regular school schedule toward the end of first semester. The past progressive and passive OPT and GJT measures were administered one week before the instruction began. GJTs were administered in regular classrooms while OPTs were administered in a quiet room with only the researcher and student present. The past progressive was the first structure to be taught. Within the same week, participants completed the aptitude test in a laboratory with the researcher and their computer teacher. Immediately after the past progressive instruction, participants were given the past progressive GJT. In the second week the passive was taught for four hours. The passive GJTs and the OPTs for both structures were administered at the end of the second week in a quiet room.

https://doi.org/10.1017/S0272263115000509 Published online by Cambridge University Press

RESULTS

Difficult Structure: The Passive

Table 1 shows Pearson correlation coefficients between the passive GJT and aptitude components (N = 62). Pearson product-moment coefficients revealed that the passive GJT pretest and posttest scores were strongly related (r *(62)* = 0.43, p < 0.01). In addition, learners' performance on the GJT posttest scores was related to LLAMA F (r *(59)*= 0.41, p < 0.01) and to LLAMA E (r *(59)* = 0.41, p < 0.01) (i.e., grammatical inferencing and sound-symbol recognition, respectively).[6]

A hierarchical multiple regression analysis was conducted, using the passive GJT posttest scores as criterion variables, and the pretest scores and LLAMA F as the predictor variables for the passive GJT scores to examine the relationships among variables. LLAMA E also significantly correlated with GJT posttest scores, but due to multicollinarity (Field, 2009), only one of the predictors had to be chosen. LLAMA F was preferred over LLAMA E because it represents the most predictive component of the language aptitude construct (i.e., grammatical inferencing). Furthermore, a recent meta-analysis conducted by Li (2015a) reported that "overall, aptitude scores and language analytical ability were more predictive than phonetic coding ability and rote-memory" (p. 403). The pretest scores were entered into the regression analysis in the first step, and then LLAMA F (grammatical inferencing) was entered in the next step. The results revealed that the pretest GJT scores significantly predicted posttest scores (β = 0. 43 t [58] = 3.62 p < 0.01). That is, learners' performance on the passive GJT pretest explained 18% of the variance (F [1, 58] = 13.11 p < 0.01) in the posttest scores. After controlling for pretest only, grammatical inferencing as measured by LLAMA F

Table 1. Pearson correlation matrix for the LLAMA subtests and passive GJT scores (N = 62)

	LLAMA B	LLAMA D	LLAMA E	LLAMA F	GJT Pretest	GJT Posttest
LLAMA B	–	0.35**	0.15	0.41**	0.08	0.18
LLAMA D		–	–0.14	0.11	–0.08	–0.02
LLAMA E			–	0.44**	0.21	0.41**
LLAMA F				–	0.15	0.41**
GJT Pretest					–	0.43**
GJT Posttest						–

** p < 0.01 (two-tailed).
Note. LLAMA B (rote memory), LLAMA D (sound recognition), LLAMA E (sound symbol correspondence), LLAMA F (grammatical inferencing).
Correlations with LLAMA subtests are based on 59 participants.

https://doi.org/10.1017/S0272263115000509 Published online by Cambridge University Press

(β = .35, t [57] = 3.18, $p < 0.05$) made a significant contribution to the regression model (see Table 2 for regression model) and grammatical inferencing ability (i.e., LLAMA F) increased the explained variance from 18% to 31% (F [2, 57] = 12.64, $p < 0.01$). That is, 12% of the variability in passive GJT posttest scores was predicted by grammatical inferencing when pretest scores were controlled.

A second multiple regression analysis was planned to examine learners' performance on the passive OPT scores in relation to the aptitude subcomponents. When correlations between aptitude components and OPT scores were screened, no significant correlations were observed among any predictor variables except for the OPT pretest and posttest scores. The passive OPT pretest scores were significantly correlated to the passive OPT posttest scores (r *(30)* = 0.46, $p < 0.01$) as expected. Correlations among the predictor and outcome variables are provided in Table 3. As indicated there are no correlations between any LLAMA subtest and passive OPT posttest scores. Based on these findings, no regression analysis was conducted because it is dependent on correlations between the criterion and predictor variables.

Easy Structure: The Past Progressive

With respect to learners' performance on the past progressive, the correlations between the past progressive GJT scores and aptitude components were examined first. Pearson product-moment coefficients revealed that pretest scores were strongly related to the past progressive GJT posttest scores (r *(62)* = 0.79, $p < 0.01$). In addition to pretest scores, LLAMA E (sound symbol recognition) was significantly related to posttest scores (r *(59)* = 0.26, $p < 0.05$). Correlations among the predictor and outcome variables are provided in Table 4.[7] Learners' scores on the

Table 2. Summary of multiple regression analysis for aptitude components predicting passive GJT scores

	B	*SE B*	β
Step 1: Control Variable			
Constant	4.27	0.66	
Passive GJT pretest	0.48	0.13	0.43**
Step 2: Aptitude Component			
Constant	3.63	0.645	
Passive GJT pretest	0.43	0.13	0.39**
LLAMA F	0.03	0	0.35**

** $p < 0.01$.
Note. R^2 = .18 for Step 1, R^2 = .31 for Step 2 ($p < 0.01$).

https://doi.org/10.1017/S0272263115000509 Published online by Cambridge University Press

Table 3. Pearson correlation matrix for the LLAMA subtests and passive OPT scores (N = 62)

	LLAMA B	LLAMA D	LLAMA E	LLAMA F	OPT Pretest	OPT Posttest
LLAMA B	–	0.35**	0.15	0.41**	0.04	0.34
LLAMA D		–	0.01	–0.14	–0.18	0.29
LLAMA E			–	0.44**	0	0.32
LLAMA F				–	0.22	0.21
Passive OPT Pretest					–	0.46**
Passive OPT Posttest						–

** $p < 0.01$ (two-tailed).
Note. LLAMA B (rote memory), LLAMA D (sound recognition), LLAMA E (sound symbol correspondence), LLAMA F (grammatical inferencing).
Correlations with OPTs are based on 30 participants.

past progressive GJT resulted in a ceiling effect indicating that students had prior knowledge of the target structure. In order to address the ceiling effect all test items were closely examined and 14 of them were removed based on the premises of classical test theory, which suggests that items that score above 80% and below 20% are not functioning well and have low discrimination power (Brown & Hudson, 2002).[8]

A hierarchical multiple regression analysis was conducted, using the past progressive GJT posttest scores as criterion variables, and the pretest scores and LLAMA E as predictor variables for the past progressive GJT scores. The pretest scores were entered into the regression analysis

Table 4. Pearson correlation matrix for the LLAMA subtests and past progressive GJT scores (N = 62)

	LLAMA B	LLAMA D	LLAMA E	LLAMA F	GJT Pretest	GJT Posttest
LLAMA B	–	0.35**	0.15	0.41	0.42**	0.2
LLAMA D		–	–0.14	0.01	0.27*	0.2
LLAMA E			–	0.44**	0.27*	0.26*
LLAMA F				–	0.38**	0.25
Passive GJT Pretest					–	0.79**
Passive GJT Posttest						–

** $p < 0.01$, * $p < 0.05$ (two-tailed).
Note. LLAMA B (rote memory), LLAMA D (sound recognition), LLAMA E (sound letter correspondence), LLAMA F (grammatical inferencing).
Correlations with LLAMA subtests are based on 59 participants.

https://doi.org/10.1017/S0272263115000509 Published online by Cambridge University Press

in the first step, and then LLAMA E was entered in the next step. The results revealed that the pretest GJT scores (β = 0.78, t [58] = 9.43 p < 0.01) significantly predicted the posttest scores. That is, learners' performance on the past progressive GJT pretest explained 61% of the variance on the posttest, (F [1, 57] = 89.06, p < 0.01). After controlling for pretest only, LLAMA E (sound-symbol correspondence) did not explain a significant proportion of variance in the past progressive GJT scores (β = 0.05, t [56] = 44.17, p > 0.05) in the regression model (see Table 5 for regression model).

Unlike the passive, the aptitude scores correlated with pretest scores but not with posttest scores for the past progressive. Given the fact that the learners had previous knowledge of the past progressive, this is not surprising. In order to address the concern for previous learning, a partial correlation was conducted where there was a strong effect of pretest on posttest in the dataset in order to tap into the unique relationship between the criterion (posttest) and possible predictors (LLAMA). Residualized posttest scores were created and new correlations were run to determine whether the residualized posttest scores correlated with LLAMA. No significant correlations were observed between any aptitude component and the residualized past progressive GJT scores (LLAMA B, r *(59)* = – 0.2, p > 0.05; LLAMA F, r *(59)* = – 0.08, p > 0.05; LLAMA D, r *(59)* = – 0.04, p > 0.05; LLAMA E, r *(59)* = 0.07, p > 0.05). This is consistent with the results of the regression model where no aptitude component was found to make a significant contribution to learners' knowledge of the past progressive. The lack of a relationship confirms that these findings are not the result of learners' prior knowledge of the target feature. The last multiple regression analysis was conducted with the past progressive OPT scores as the outcome variable. In order to decide which variables might be the best predictors for the regression analysis to be conducted on the past progressive OPT posttest, correlations between the dependent variable and aptitude components

Table 5. Summary of multiple regression analysis for aptitude components predicting past progressive GJT Scores

	B	*SE B*	*β*
Step 1: Control Variable			
Constant	0.47	0.19	
Past Progressive GJT pretest	0.94	0.1	0.78**
Step 2: Aptitude Component			
Constant	0.44	0.2	
Past Progressive GJT pretest	0.92	0.1	0.77**
LLAMA E	0	0	0.05

** p < 0.01.
Note. R^2 = .61 for Step 1, R^2 = .61 for Step 2 (p < 0.01).

https://doi.org/10.1017/S0272263115000509 Published online by Cambridge University Press

were checked (see Table 6 for correlations among predictors and the past progressive OPT scores). Pearson product-moment coefficients revealed that the progressive OPT posttest (dependent variable) was highly correlated with OPT pretest scores (r *(30)* = 0.75, $p < 0.01$) and also related to LLAMA B, that is, vocabulary learning (r *(30)* = 0.51, $p < 0.01$). Next, the OPT pretest scores were entered into regression analysis, followed by LLAMA B.

Results of the multiple regression analysis revealed that the progressive pretest OPT scores ($\beta = 0.75$, t [28] = 5.93, $p < .001$) significantly predicted the posttest OPT scores. Table 7 summarizes the regression model. Progressive OPT pretest scores explained 55% of the variance in the posttest scores (F [1, 28] = 35.21, $p < 0.01$). After controlling for OPT pretest scores, LLAMA B (i.e., vocabulary learning) ($\beta = 0.27$, t [27] = 2.1, $p < 0.05$) explained a significant proportion of variance in OPT posttest scores and explained the increased variance from 56% to 62%, (F [2, 27] = 21.91, $p < 0.01$). The results revealed that there was a significant R square change (6%) explained by LLAMA B when pretest scores were controlled.

DISCUSSION

The research reported here examined the unique contributions of aptitude and its subcomponents to the learning of two target structures defined as easy or difficult to learn. The results on the GJT for the passive (difficult structure) indicate that when learners were judging the passive items they relied more on their language aptitude, particularly grammatical inferencing ability. Grammatical inferencing is known to

Table 6. Pearson correlation matrix for the LLAMA subtests and past progressive OPT scores (N = 62)

	LLAMA B	LLAMA D	LLAMA E	LLAMA F	OPT Pretest	OPT Posttest
LLAMA B	–	0.35**	0.15	0.41**	0.38*	0.51**
LLAMA D		–	–0.14	0.01	0.03	0.15
LLAMA E			–	0.44**	0.07	–0.06
LLAMA F				–	0.15	0.32
Past Progressive OPT Pretest					–	0.75**
Past Progressive OPT Posttest						–

* $p < 0.05$; ** $p < 0.01$ (two-tailed).
Note. LLAMA B (rote memory), LLAMA D (sound recognition), LLAMA E (sound symbol correspondence), LLAMA F (grammatical inferencing).
Correlations with OPTs are based on 30 participants.

https://doi.org/10.1017/S0272263115000509 Published online by Cambridge University Press

Table 7. Summary of multiple regression analysis for aptitude components predicting past progressive OPT scores

	B	*SEB*	*β*
Step 1: Control Variable			
Constant	155.45	61.49	
Past Progressive OPT pretest	0.78	0.13	0.75**
Step 2: Aptitude Component			
Constant	–29.04	105.62	
Past Progressive OPT pretest	0.68	0.13	0.65**
LLAMA B	4.42	2.11	0.27*

** $p = 0.01$, *$p = 0.05$.
Note. $R^2 = .56$ for Step 1, $R^2 = .62$ for Step 2 ($p < .001$).

be the most predictive and representative subtest among all aptitude components (Li, 2015a). The same aptitude subcomponent was not a significant factor affecting learners' performance on the past progressive GJT. This is consistent with our hypothesis that when the participants were taught the more difficult structure (i.e., the passive) they had to rely on their language aptitude to fully understand a language feature that is formally complex, opaque in form-meaning connections, and relatively low in terms of salience in the input. This is consistent with Li's (2013) hypothesis referred to in the preceding text that both high aptitude and explicit conditions are necessary for learning difficult structures.

The explicit instruction that learners received about the passive probably served to initiate their awareness of it. Subsequently learners with greater grammatical inferencing ability likely continued to formulate hypotheses and to look for regularities in the input to confirm their hypotheses about the passive. Learners with lower levels of grammatical inferencing ability probably did not engage in this process to the same extent. Thus, grammatical inferencing ability appears to have served as an additional resource for learning a difficult structure with a group of learners who all received explicit instruction and opportunities to practice it in communicative activities. It should also be remembered that the passive was a novel structure for the learners. Thus they were at an early stage of development with respect to this form. According to Skehan's (2002) aptitude profile model, learners at the *patterning* stages of development may derive particular benefits from the grammatical inferencing component of aptitude. The fact that LLAMA F (grammatical inferencing) predicted learners' performance is consistent with Skehan's model and Li's (2015a) findings in his meta-analysis.

Different aptitude components appear to have been used as a cognitive resource with respect to learners' performance on the past

https://doi.org/10.1017/S0272263115000509 Published online by Cambridge University Press

progressive tasks. That is, *rote (associative) memory* ability, as measured by the *vocabulary-learning* subtask (i.e., LLAMA B), played a role when it came to learners' use of the past progressive on the oral production task but there was no role for aptitude on the written GJT. It is important to recall that this target structure was not novel to the students and therefore they already possessed some explicit knowledge of this form prior to the instruction. Nonetheless, learners improved significantly from pretest to posttest. Thus it is possible that the learning of the past progressive was less related to aptitude because it is structurally simple, highly salient, and frequent in the input. This may have attenuated the contribution of aptitude in developing explicit knowledge of this structure.[9] Given the lack of correlation between passive GJT pretest scores and aptitude scores, one might claim that the role of aptitude is more evident in the development of explicit knowledge of the passive in comparison to past progressive. It seems that previous learning levels out or at least minimizes the role of aptitude. Hence novelty may be a factor that mediates the role of aptitude in this dataset. Nonetheless it is important to keep in mind that if the past progressive had not been taught earlier the results might have been different. It is also worthwhile to remember that Li (2015a) in his meta-analysis reports relatively low effect sizes for the role of aptitude in comparison to earlier studies but he also acknowledges the complex interplay of many factors in SLA and states that "SLA is subject to multitude of learner-internal and learner-external factors, the amount of variance (around 10 per cent) accounted for by aptitude should by no means be taken as small" (p. 404).

The results of this study appear to offer some support for the aptitude profile model proposed by Skehan (2002). For example, learners with higher scores on LLAMA B (the vocabulary learning test taken as a measure of memory) performed significantly better on the past progressive OPT than learners with lower scores. This is consistent with the prediction of the aptitude profile model that memory plays a greater role at later stages in the development of a structure (the past progressive having already been acquired to some extent by the onset of the study). The differences in the findings with regard to learners' written and oral performance on both structures also provide some support for the model. That is, when learners encountered a difficult/novel structure (i.e., the passive) they worked through *patterning* using their LAA. A close examination of the GJT scores showed that learners were able to identify the errors and to correct them, suggesting that after going through the *pattern identification* stage they were able to make generalizations based on the regularities in the input they had received. However, because the instruction only lasted for four hours, they probably did not have enough time to automatize this knowledge. Therefore, they were not successful in extending their understanding of the passive to use it in oral production. This may have resulted in a lack of a relationship between language

https://doi.org/10.1017/S0272263115000509 Published online by Cambridge University Press

aptitude and oral production performance for the passive. *Automatizing* and *lexicalising* are the last two steps in Skehan's model. Arguably, learners in this study needed more time to reach and move through these stages with respect to a target structure that is more difficult.

As indicated in the preceding text, the aptitude component that contributed most to learners' performance on the past progressive OPT was memory. Learners were at later stages of processing (i.e., *lexicalizing*) with respect to the past progressive. According to Skehan's model, (2002) learners at later stages of processing are "able to choose freely, and appropriately, between lexicalized representation of an interlanguage form, and its rule-based creative version" (p. 89). It was evident that in oral production learners were producing the past progressive frequently. They were also able to switch between *when* and *while* clauses in the past progressive. This is in line with rule-based creative use suggested by Skehan. The lack of relation between grammatical inferencing and oral production scores, which represent more spontaneous use of language, is not necessarily out of line as earlier research has reported similar findings with fluency (e.g., Harley & Hart, 1997; Ranta, 2005).

The findings of the study are in line with Carroll's model of school learning, which posits that language aptitude is only one of the factors contributing to L2 success. He specifies *time* and *instructional excellence* as two other main factors. The ATI studies reviewed earlier provide compelling evidence that some features of L2 instruction can minimize the effects of language aptitude. Erlam (2005) reported that one type of L2 instruction was resistant to effects of aptitude, and she explained this finding in relation to the higher levels of motivation that resulted from the match between learner preferences and type of L2 instruction. According to Erlam this match—instructional excellence in Carroll's Model of School learning—resulted in minimizing the effects of aptitude. In the present study all learners were provided with meaningful input and concise explicit information integrated into meaningful practice for two structures. Yet only the learning of one of the structures appears to have been influenced by aptitude. This suggests that the nature of the target structure interacts with type of L2 instruction and language aptitude. In this study the instructional input was the same for both language features, but it seems that aptitude worked as a facilitator for the difficult structure (passive) while it had a minimal effect for the less difficult structure (past progressive).

As with all research there are limitations to this study including the fact that one of the features investigated was novel to the learners and the other was not. In future research it will be important to select two structures that are both novel[10] but differ in terms of difficulty. Undoubtedly this is easier in a laboratory than in a classroom setting, particularly with intact classes and a set curriculum (for an example of lab study, see Godfroid, this issue). It is important to point out that differentiating

https://doi.org/10.1017/S0272263115000509 Published online by Cambridge University Press

language structures in terms of difficult or easy is complicated. Thus the criteria used to distinguish the two structures in this study could be challenged based on different conceptual/theoretical perspectives. Other limitations include the absence of a control group and the short duration of the instructional treatment. It is also important to keep in mind that this study was conducted in an EFL context with learners who had received several years of structure-based English language instruction. Thus, the findings may not be applicable to other L2 learner populations.

CONCLUSION

The findings demonstrate that different aptitude components contribute to the learning of difficult and easy L2 structures in different ways. The findings also provide some support for Skehan's aptitude profile model linking different components of aptitude to different stages of language acquisition (i.e., earlier and later stages of L2 development). The findings underscore the complex interplay among several factors in the classroom setting. DeKeyser (2012) states that "aptitude is rendered crucial by the demands of the learning processes encouraged (allowed?) by a specific instructional treatment" (p. 192). The fact that the same type of instruction provided for both structures was enough for learners to significantly progress in their knowledge of the past progressive but not the passive without language aptitude as a resource demonstrates that the picture is not complete without consideration of the role of the target structure.

As indicated in the introduction there is only one laboratory study that has investigated the role of aptitude in relation to the learning of different language structures (Robinson, 2002) and, to our knowledge, the study reported here is the only classroom one. Given the potential that such research has for a better understanding of how different L2 structures are learned and the implications for L2 teaching, there is a need for future instructed SLA studies to focus on this issue in both classroom and laboratory settings.

SUPPLEMENTARY MATERIAL

To view supplementary material for this article, please visit http://dx.doi.org/10.1017/S0272263115000509.

Received 30 March 2015
Accepted 11 November 2015
Final Version Received 17 November 2015

https://doi.org/10.1017/S0272263115000509 Published online by Cambridge University Press

NOTES

1. There are other conceptualizations of memory apart from those discussed in relation to the MLAT or other language aptitude tests. E.g., two long-term memory systems referred to as declarative and procedural are considered to contribute to the representation of knowledge in L1 and L2 learning (Ullman, 2004).

2. For more detailed discussions of distinctions between difficulty and complexity see Bulté & Housen (2012), Housen (2014), and Robinson (2002).

3. Private schools offer more English language classes for students. The regular curriculum includes four hours of English at eight grade but at private schools this can go up to seven to eight hours.

4. While controlling for verb type (i.e., telic and activity verbs) in both grammatical and ungrammatical items, and error type in ungrammatical items in past progressive it was not possible to have even number of grammatical and ungrammatical items.

5. It is important to emphasize that instances of L2 performance are best viewed not as relying on either implicit or explicit knowledge but, rather, on both types of L2 knowledge to varying degrees. Thus we predict that learners' performance on the OPT is a reflection of greater access to implicit L2 knowledge and on the GJT to greater access to explicit L2 knowledge.

6. However, it should be noted that there is enough variance in the passive GJT pretest scores to be the basis for correlations in the posttest scores.

7. Correlations and regression analyses were conducted with the edited version of the past progressive GJTs with 26 target items.

8. It is of common practice to run item and reliability analyses in any test and to ensure basic test qualities before any conclusion about the test scores can be drawn. In our case, with the removal of 14 items stated in the preceding text we achieved a higher reliability score and therefore more consistency in the test. The reliability coefficient of the past progressive GJT (pretest) was .95 before the changes in the test. Then it was .96 after the revisions in the number of items. Reliability of the posttest remained the same .96 after the revision.

9. The findings do not diminish the pretest effects, but the post-hoc analyses to control for the shared variance between pretest and posttest performance on the written GJT confirmed no relationship between aptitude and learner's knowledge of the past progressive.

10. Regular school curriculum and EFL course books over L2 structures more or less in a fixed order in more EFL contexts. Past progressive is introduced before the passive construction. Therefore, inevitably in a classroom-based research, passive would be a novel structure whereas past progressive would be familiar one at any given point for an intervention study.

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.

Brown, J. D., & Hudson, T. (2002). *Criterion-referenced language testing*. Cambridge, UK: Cambridge University Press.

Bulté, B., & Housen, A. (2012). Defining and operationalising L2 complexity. In A. Housen, F. Kuiken, & I. Vedder (Eds.). *Dimensions of L2 performance and proficiency: Complexity, accuracy and fluency in SLA* (pp. 21–46). Amsterdam, The Netherlands: John Benjamins.

Carroll, J. B. (1981). Twenty-five years of research in foreign language aptitude. In K. C. Diller (Ed.), *Individual differences and universals in language learning aptitude* (pp. 83–113). Rowley, MA: Newbury House.

Carroll, J. B. (1990). Estimating item and ability parameters in homogeneous tests with the person characteristic function. *Applied Psychological Measurement*, *14*, 109–125.

Carroll, J. B., & Sapon, S. (1959). *The modern languages aptitude test*. San Antonio, TX: Psychological Measurement.

https://doi.org/10.1017/S0272263115000509 Published online by Cambridge University Press

Collins, L., Trofimovich, P., White, J., Cardoso, W., & Horst, M. (2009). Some input on the easy/difficult grammar question: An empirical study. *The Modern Language Journal*, *93*, 336–353.

de Graaff, R. (1997). The eXperanto experiment: Effects of explicit instruction on second language acquisition. *Studies in Second Language Acquisition*, *19*, 249–275.

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

DeKeyser, R. M. (2000). The robustness of critical period effects in second language acquisition. *Studies in Second Language Acquisition*, *22*, 509–544.

DeKeyser, R. M. (2005). What makes learning second-grammar difficult? A review of issues. *Language Learning*, *55* (suppl.), 1–25.

DeKeyser, R. M. (2012). Interactions between individual differences, treatments, and structures in SLA. *Language Learning*, *62* (suppl. 2), 189–200.

Doughty, C., & Williams, J. (1998). Issues and terminology. In C. Doughty and J. Williams (Eds.), *Focus on form in classroom second language acquisition* (pp. 1–11). Cambridge, UK: Cambridge University Press.

Dörnyei, Z. (2005). *The psychology of the language learner: Individual differences in second language acquisition*. Mahwah, NJ: Lawrence Erlbaum.

Dörnyei, Z., & Skehan, P. (2003). Individual differences in second language learning. In C. Doughty & M. H. Long (Eds.), *The handbook of second language acquisition* (pp. 589–630). Oxford, UK: Blackwell.

Ehrman, M. E., & Oxford, R. (1995). Cognition plus: Correlates of language learning success. *The Modern Language Journal*, *79*, 67–89.

Ellis, R. (2009). Measuring implicit and explicit knowledge of a second language. In R. Ellis, S. Loewen, C. Elder, R. Erlam, J. Philp, and H. Reinders (Eds.), *Implicit and explicit knowledge in second language learning, testing and teaching* (pp. 27–30). Bristol, UK: Multilingual Matters.

Erlam, R. (2005). Language aptitude and its relationship to instructional effectiveness in second language acquisition. *Language Teaching Research*, *9*, 147–171.

Field, A. P. (2009). *Discovering statistics using SPSS Statistics*, 3rd ed. London: Sage.

Goldschneider, J. M., & DeKeyser, R. M. (2001). Explaining the "Natural Order of L2 Morpheme Acquisition" in English: A meta-analysis of multiple determinants. *Language Learning*, *51*, 1–50.

Granena, G. (2013). Cognitive aptitudes for second language learning and the LLAMA language aptitude test. In G. Gisela & M. Long (Eds.), *Sensitive periods, language aptitude, and ultimate attainment* (pp. 105–130). Amsterdam, The Netherlands: John Benjamins.

Granena, G., & Long, M. (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

Grigorenko, E. L., Sternberg, R. J., & Ehrman, M. E. (2000). A theory based approach to the measurement of foreign language learning ability: The Canal-F theory and test. *Modern Language Journal*, *84*, 390–405.

Harley, B., & Hart, D. (1997). Language aptitude and second language proficiency in classroom learners of different starting ages. *Studies in Second Language Acquisition*, *19*, 379–400.

Harley, B., & Hart, D. (2002). Age, aptitude and second language learning on bilingual exchange. In P. Robinson (Ed.), *Individual differences and instructed language learning* (pp. 301–330). Amsterdam, The Netherlands, and Philadelphia, PA: John Benjamins.

Harrington, M., & Sawyer, M. (1992). L2 working memory capacity and L2 reading skill. *Studies in Second Language Acquisition*, *14*, 25–38. doi: 10.1017/S0272263100010457

Hinkel, E. (2004). Tense, aspect and the passive voice in L1 and L2 academic texts. *Language Teaching Research*, *8*, 5–29.

Housen, A. (2014). Difficulty and complexity of language features and second language instruction. *The encyclopedia of applied linguistics*, 1–7. Oxford, UK: Wiley-Blackwell.

Housen, A., Pierrard, M., & Van Daele, S. (2005). Structure complexity and the efficacy of explicit grammar instruction. In A. Housen & M. Pierrard (Eds.), *Investigations in instructed second language acquisition* (pp. 235–269). Brussels: Mouton DeGruyter.

https://doi.org/10.1017/S0272263115000509 Published online by Cambridge University Press

Hulstijn, J., & de Graaff, R. (1994). Under what conditions does explicit knowledge of a second language facilitate the acquisition of implicit knowledge? A research proposal. *AILA Review*, *11*, 97–112.
Hwu, F., & Sun, S. (2012). The aptitude-treatment interaction effects on the learning of grammar rules. *System*, *40*, 505–521.
Kirby, S. (2010). Passives in first language acquisition: What causes the delay? *University of Pennsylvania Working Papers in Linguistics*, *16*, 109–117.
Kormos, J., & Sáfár, A. (2008). Phonological short-term memory and foreign language performance in intensive language learning. *Bilingualism: Language and cognition*, *11*, 261–171.
Li, S. (2013). The interactions between the effects of implicit and explicit feedback and individual differences in language analytic ability and working memory. *Modern Language Journal*, *97*, 634–654.
Li, S. (2015a). The associations between language aptitude and second language grammar acquisition: A meta-analytic review of five decades of research. *Applied Linguistics*, *36*, 385–408.
Li, S. (2015b). The differential roles of two aptitude components in mediating the effects of two types of feedback on the acquisition of an opaque linguistic structure. In C. Sanz and B. Lado (Eds.) *Individual differences, L2 development, and language program administration: From theory to application* (pp. 32–52). Boston, MA: Cengage Learning.
Linck, J. A., Hughes, M. M., Campbell, S. G., Silbert, N. H., Tare, M., Jackson, S. R., ... and Doughty, C. J. (2013). Hi-LAB: A new measure of aptitude for high-level language proficiency. *Language Learning*, *63*, 530–566.
Meara, P. (2005). *LLAMA language aptitude tests*. Swansea: Lognostics.
Miyake, A., & Friedman, N. P. (1998). Individual differences in second language proficiency: Working memory as language aptitude. In A. Healy, & L. Bourne (Eds.), *Foreign language learning: Psycholinguistics studies on training and retention* (pp. 339–364). Mahwah, NJ: Lawrence Erlbaum.
Osaka, M., & Osaka, N. (1992). Language-independent working memory as measured by Japanese and English reading span tests. *Bulletin of the Psychonomic Society*, *30*, 287–289.
Pimsleur, P. (1966). *The Pimsleur language aptitude battery*. New York, NY: Harcourt Brace Jovanovich.
Ranta, L. (2005). Learners' analytic ability and oral production in a second language: Is there a connection? In A. Housen & M. Pierrard (Eds.), *Investigations in instructed second language acquisition* (pp. 99–130). Berlin: Mouton de Gruyter.
Révész, A. (2007). Focus on form in task-based language teaching: Recasts, task complexity, and L2 learning (Unpublished doctoral dissertation). Columbia: Teachers College.
Révész, A. (2009). Task complexity, focus on form, and second language development. *Studies in Second Language Acquisition*, *31*, 437–470. Columbia University, New York, USA.
Robinson, P. (2001). Individual differences, cognitive abilities, aptitude complexes and learning conditions in second language acquisition. *Second Language Research*, *17*, 368–392.
Robinson, P. (2002). Effects of individual differences in intelligence, aptitude, and working memory on adult incidental SLA: A replication and extension of Reber, Walkenfield, and Hernstad (1991). In P. Robinson (Ed.), *Individual differences and instructed language learning* (pp. 211–265). Amsterdam, The Netherlands: John Benjamins.
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). Oxford, UK: Springer.
Rysiewicz, J. (2008), Cognitive profiles of (un)successful FL learners: A cluster analytical study. *The Modern Language Journal*, *92*, 87–99. doi: 10.1111/j.1540-4781.2008.00688.x
Sasaki, M. (1996). *Second language proficiency, foreign language aptitude, and intelligence*. New York, NY: Lang.
Sheen, Y. (2007a). The effect of focused written corrective feedback and language aptitude on ESL learners' acquisition of articles. *TESOL Quarterly*, *41*, 255–283.

https://doi.org/10.1017/S0272263115000509 Published online by Cambridge University Press

Sheen, Y. (2007b). The effect of corrective feedback, language aptitude and learner attitudes on the acquisition of English articles, In A. Mackey (Ed.), *Conversational interaction in second language acquisition* (pp. 301–322). Oxford, UK: Oxford University Press.

Schmidt, R. (1995). Consciousness and foreign language learning: A tutorial on the role of attention and awareness in learning. In R. Schmidt (Ed.), *Attention and awareness in foreign language learning* (pp. 1–63). Honolulu: University of Hawaii, Second Language Teaching & Curriculum Center.

Skehan, P. (1986). The role of foreign language aptitude in a model of school learning. *Language Testing*, *3*, 188–221.

Skehan, P. (2002). Theorising and updating aptitude. In P. Robinson. (Ed.), *Individual differences and instructed language learning* (pp. 69–93). Amsterdam, The Netherlands: John Benjamins.

Skehan, P. (2010). Language aptitude. In S. Gass & A. Mackey (Eds.), *Handbook of Second Language Acquisition* (pp. 381–395). London: Routledge.

Skehan, P. (2015). Foreign language aptitude and its relationship with grammar: A critical overview. *Applied Linguistics*, *36*, 367–384.

Spada, N. (1997). Form-focussed instruction and second language acquisition: A review of classroom and laboratory research. *Language Teaching*, *30*, 73–87.

Spada, N., & Tomita, Y. (2010). Interactions between type of instruction and type of language feature: A meta-analysis. *Language Learning*, *60*, 263–308.

Spada, N., Jessop, L., Suzuki, W., Tomita, Y., & Valeo, A. (2014). Isolated and integrated form-focused instruction: Effects on different types of L2 knowledge. *Language Teaching Research*, *18*, 453–473.

Sparks, R. L., Patton, J., Ganschow, L., & Humbach, N. (2011). Subcomponents of second-language aptitude and second-language proficiency. *The Modern Language Journal*, *95*, 253–273.

Sternberg, R. J. (1999). Theory of successful intelligence. *Review of General Psychology*, *4*, 292–316.

Swan, M. (1995). *Practical English usage*. Oxford, UK: Oxford University Press.

Trofimovich, P., Ammar, A., & Gatbonton, E. (2007). How effective are recasts? The role of attention, memory, and analytical ability. In A. Mackey (Ed.), *Conversational interaction in second language acquisition: A series of empirical studies* (pp. 171–195). Oxford, UK: Oxford University Press.

Ullman, M. T. (2004). Contributions of memory circuits to language: The declarative/procedural model. *Cognition*, *92*, 231–270.

Yalçın, Ş. (2012). Individual differences and the learning of two grammatical features with Turkish learners of English. Unpublished doctoral dissertation, University of Toronto, Canada.

Yalçın, Ş., Çeçen, S., & Erçetin, G. (forthcoming). The relationship between aptitude and working memory: An instructed SLA context. Manuscript submitted for publication.

Yilmaz, Y. (2012). Relative effects of explicit and implicit feedback: The role of working memory capacity and language analytical ability. *Applied Linguistics*, *34*, 344–368.

Wesche, M. (1981). Language aptitude measures in streaming, matching students with methods, and diagnosis of learning problems. In K. C. Diller (Ed.), *Individual differences and universals in language learning aptitude* (pp. 119–154). Rowley, MA: Newbury House.

VanPatten, B., & Smith, M. (2015). Aptitude as grammatical sensitivity and the initial stages of learning Japanese as a L2: Parametric variation and case marking—corrigendum. *Studies in Second Language Acquisition*, *37*, 175–176.

VanPatten, B., Collopy, E., Price, J. E., Borst, S. & Qualin, A. (2013). Explicit information, grammatical sensitivity, and the first-noun principle: A cross-linguistic study in processing instruction. *The Modern Language Journal*, *97*, 506–527.

https://doi.org/10.1017/S0272263115000509 Published online by Cambridge University Press