

ARTICLE

# The associations between cognitive ability and L2 development under five different instructional conditions

Shaofeng Li<sup>1\*</sup>, Rod Ellis<sup>2</sup> and Yan Zhu<sup>3</sup>

<sup>1</sup>Florida State University and Guangdong University of Foreign Studies, <sup>2</sup>Curtin University and <sup>3</sup>Fudan University

\*Corresponding author. Email: [sli9@fsu.edu](mailto:sli9@fsu.edu)

(Received 16 June 2018; revised 03 November 2018; accepted 07 November 2018)

## Abstract

This study examines the effects of two cognitive abilities—language analytic ability (LAA) and working memory (WM)—on language learning under five different instructional conditions. One hundred fifty eighth-grade English as a foreign language learners underwent a 2-hr treatment session. They were divided into five groups based on whether and when they received form-focused instruction. One group received pretask instruction on the linguistic target (English passive voice) before performing two narrative tasks; a second group received within-task feedback but no pretask instruction; a third group received both pretask instruction and within-task feedback; a fourth group received feedback after completing the tasks; and the fifth group only performed the tasks. The results showed that (a) LAA was predictive of the posttest scores of the group that only performed the communicative tasks and the group who received posttask feedback, (b) WM was associated with the learning outcomes of the two groups receiving within-task feedback, and (c) neither cognitive variable was implicated in the group that received pretask instruction before performing the tasks. The results suggest that the impact of LAA is evident when there is less external assistance and that WM is involved when learners face the heavy processing burden imposed by within-task feedback.

**Keywords:** language analytic ability; language aptitude; second language acquisition; task-based language instruction; working memory

Language aptitude is a set of cognitive abilities that have been shown to strongly predict and explain the process and product of language learning. Two broad streams of aptitude research have emerged: predictive and interactional (Li, 2015a, 2016, 2017a, 2018). In the predictive approach, aptitude is viewed as a trait variable that impacts on the outcome of learning, regardless of context and instruction type. The objectives of such research are to (a) ascertain the predictive power of aptitude in itself and in comparison with other variables, (b) identify those learners who will succeed in a language program, and (c) identify individuals with learning disabilities to waive their language

requirements or provide them with additional support (Carroll & Sapon, 2002). In the interactional approach, the focus is on whether there is a fit between learners' aptitude strengths and the processing demands of the learning condition. Interactional research investigates whether the same cognitive variable has differential effects under different treatment conditions and/or whether different cognitive variables mediate the effects of the same treatment in different ways. Interactional research is of greater theoretical and pedagogical significance than predictive research because it can show not only the joint and unique effects of aptitude and treatment on learning outcomes but also the processes linking the two (DeKeyser, 2012). However, as DeKeyser observed, there is a lack of interactional research where variables are strictly manipulated and controlled and treatments are consistently implemented. The study reported in this article employs an interactional approach by investigating whether two aptitude components—language analytic ability and working memory—are drawn upon differently by five treatment groups that differ in terms of whether they received form-focused instruction and whether they received it before, during, or after performing communicative tasks.

Language analytic ability and working memory have become a favorite duo for researchers interested in the interplay between cognitive ability and form-focused instruction (e.g., Suzuki & DeKeyser, 2016; Yilmaz, 2013) because of their importance in influencing learning success and because of the different mechanisms through which they affect learning. Language analytic ability is essential for rule extrapolation and is implicated in online or offline tasks where learners are engaged in searching for rules or processing meta-linguistic information (Skehan, 2012). Working memory is a cognitive space for simultaneous storage and processing of linguistic input and is implicated in tasks where learners face online processing pressure (Li, 2017b). Working memory is also assumed to be important for noticing linguistic input in the currently popular cognitive–interactionist approach to language learning (Long, 2016). Although working memory involves both information storage and processing, the processing component is a domain-general cognitive device responsible for attention control, which is fundamentally different from domain-specific language analytic ability (Li, 2017a). Therefore, these two cognitive variables are theoretically distinct and are purported to make unique contributions to language learning. Our interest is in investigating whether and in what way these two variables are associated with the effects of different types of form-focused instruction integrated with meaning-oriented tasks.

Form-focused instruction is defined as any attempt to draw learners' attention to linguistic forms in communicative classes (Spada, 2011). There are different theoretical positions over the ideal time for, and the necessity of, addressing linguistic forms in a communicative class. According to skill acquisition theory (DeKeyser, 2015), learners must first have a solid base of declarative knowledge before engaging in skill-specific practice activities for proceduralization and automatization to occur. This theory supports a PPP (present–practice–produce) approach where learners receive explicit instruction and controlled practice before engaging in free production tasks. Providing explicit instruction before task performance has been labeled as “task-supported teaching” by Ellis (2003, p. 28) and is seen as “a weak form” of task-based language teaching by Skehan (1996, p. 39). However, Long (2015, 2016) has consistently rejected pretask grammar instruction on the grounds that learners

follow their internal timetable in mastering grammatical morphemes and may therefore not be ready to acquire a preselected target structure. He contends that focus on form must be reactive, incidental, and contingent, and should happen “during (and if necessary after, but not before) task work” (2016, p. 17). Willis and Willis (2007) also oppose teaching grammar in the pretask stage because the learner will focus on the linguistic target, which will affect fluency and undermine the meaning-primary nature of the task. Willis and Willis suggest postponing grammar instruction to the posttask stage where there is no risk of a negative influence on fluency or meaning making. Finally, according to Krashen’s (1981) well-known input theory, language learning occurs primarily through exposure to and use of the second language, and there is no need to attend to linguistic forms.

What is missing in the theoretical debate over the necessity and timing of form-focused instruction, however, is how the various options affect the way learners deploy their cognitive resources. Our study seeks to explore the interface between the options of form-focused instruction outlined above and two important cognitive factors—language analytic ability and working memory—with a view to providing empirical evidence for theory building and pedagogical decision making. To the best of our knowledge, no study has examined the effects of different cognitive factors on language learning resulting from such a range of form-focused instruction options. In addition, the study overcomes the limitations of previous studies by using validated tests for the predictor and criterion variables, implementing the treatments in a classroom setting to strive for ecological/external validity, and including learners’ pretest scores as a predictor to show the impact of cognitive ability after controlling for learners’ existing knowledge of the target structure. In what follows, we review the empirical studies examining the associations between the two cognitive variables and second language grammar learning. Given that we examine aptitude–treatment interaction rather than whether cognitive ability is predictive of ultimate learning outcome, our review will focus on the interactional studies, especially those most relevant to our research questions.

### Language analytic ability

Language analytic ability refers to (a) the ability to understand the grammatical functions of sentence elements, which is measurable via the grammatical sensitivity subtest of the Modern Language Aptitude Test (MLAT; Carroll & Sapon, 2002), and (b) the ability to extrapolate linguistic regularities from input materials, which can be gauged using LLAMA-F (Meara, 2005) or the language analysis subtest of the Pimsleur Language Aptitude Battery (PLAB; Pimsleur, Reed, & Stansfield, 2004). Li’s (2015a) meta-analysis of the results of 25 studies showed that language analytic ability was significantly correlated with L2 grammar learning,  $r = .35$ , 95% confidence interval [.27, .43], which is considered to be a near-medium effect based on Plonsky and Oswald’s (2014) criteria for assessing the magnitude of an effect size. The meta-analysis further demonstrated that the correlations between aptitude and learning outcomes were moderated by methodological factors. The predictive research indicated that analytic ability had a stronger predictive power for high school students than for university students, suggesting that aptitude is likely more

relevant for initial L2 learning than for more advanced learning. For the interactional studies, Li coded the instructional treatments into implicit and explicit according to whether the treatment overtly drew learners' attention to linguistic forms. Explicit treatments were found to be more strongly correlated with analytic ability than implicit treatments,  $r = .40$  versus  $r = .17$ . However, while the meta-analysis contributes to our understanding of the construct of aptitude, it also masks the idiosyncrasies of the primary studies and the methodological variation among them. In the following sections, a more detailed analysis of the treatment studies related to this study is provided.

One strand of interactional research in this domain has revolved around whether analytic ability is correlated with the effects of corrective feedback, a variable examined in this study. Overall, the research (Sachs, 2010; Sheen, 2007; Yilmaz, 2013) indicated that language analytic ability was correlated with the effects of explicit feedback, consisting of metalinguistic correction or explicit correction, but not those of implicit feedback, operationalized as recasts reformulating learners' erroneous utterances or as signals indicating whether learners had correctly understood linguistic descriptions of some photos. It was also not implicated when learners simply engaged in meaning-processing without receiving feedback. The researchers' explanation of these results is that a prerequisite for the role of analytic ability is awareness, which is argued to be present in explicit instruction but absent in implicit instruction.

While it might be a catalyst for aptitude effects, awareness is a complex cognitive behavior that is subject to a variety of learner-external and learner-internal factors such as the instructional context, the saliency of the linguistic target, learners' previous experience, learners' age, and the research setting, to name a few. For example, learners in intensive language programs are more likely to treat a communicative task as an activity for language practice than those in immersion programs or content-based classes (Ellis, 2003), and salient structures are more likely to be noticed than nonsalient structures (Mackey, Gass, and McDonough, 2000). Therefore, awareness cannot be attributed only to the explicitness/implicitness of the instructional treatment, although researchers may be able to manipulate the design features of the treatment to achieve the desired effect. The subtle nature of awareness may account for the conflicting findings of existing research. For example, contrary to other studies, Trofimovich, Ammar, and Gatonton (2007) reported that analytic ability predicted the effects of recasts, perhaps because the intensive, computerized recasts provided in a lab setting were notably salient and thus likely enhanced learners' awareness.

Another factor that mediates the relationship between analytic ability and treatment type is structural difficulty/complexity. Li (2013a, 2013b) found that analytic ability was a significant predictor of the effects of metalinguistic feedback for a complex structure, but not for a simple structure. Yalçın and Spada (2016) reported that after receiving form-focused instruction consisting of rule explanation and communicative practice (the authors did not describe what tasks were used), analytic ability was a significant predictor for the learning of the English passive voice—a difficult structure—but not of the past progressive—an easy structure. However, Robinson (1997) reported that under a rule search condition where learners were encouraged to look for rules based on given sentences, analytic ability was correlated with the

learning of an easy rule—English subject–verb inversion—but not that of a hard rule—the formation of English pseudoclefts. These studies suggest that when the linguistic structure is easy, the effect of language analytic ability is evident in the absence of rule explanation, but that when rule explanation is available, the role of analytic ability is neutralized. When the structure is difficult, its effects surface if there is rule explanation (because understanding the difficult rule requires analytic ability) but not if there is no rule explanation as then it is beyond learners' processing capacity.

Two studies have examined how analytic ability interfaces with deductive and inductive instruction. An often-cited study by Erlam (2005) showed that analytic ability was uncorrelated with the effects of deductive instruction where learners were given grammar instruction followed by form-focused production activities, but it was related to the effects of inductive instruction where learners were prompted to make inferences about the grammar rule when performing similar activities. The study also showed a significant correlation between analytic ability and the effects of structured input consisting of grammar explanation followed by comprehension activities. Erlam speculated that the grammar instruction together with production activities in the deductive treatment neutralized the influence of analytic ability. Hwu and Sun (2012) and Hwu, Wei, and Sun (2014) explored whether high- and low-aptitude learners benefited from deductive and inductive instruction in different ways. They found that deductive instruction was more effective for low-aptitude learners while inductive instruction was more effective for high-aptitude learners. These two studies suggest that providing explicit information about the target structure may neutralize differences in aptitude and so benefit low-aptitude learners, but it may disadvantage high-aptitude learners who learn more from inductive instruction.

To sum up, the predictive research shows that language analytic ability has an overall strong correlation with grammar learning. However, the interactional research shows that its associations with learning outcomes vary depending on the nature of the instruction: it is more strongly correlated with explicit instruction than implicit instruction. When the instruction is explicit, its role varies depending on whether the linguistic target is easy or difficult and on whether the instruction requires deductive or inductive learning. Although these studies have not directly examined the associations between analytic ability and the timing and presence/absence of form-focused instruction, which are the foci of this study, they provide useful clues we can draw on to interpret our results and make predictions about our findings. For instance, performing communicative tasks without form-focused instruction, a treatment condition investigated in this study, constitutes an implicit learning condition that may not show strong correlations with analytic ability, as the literature has demonstrated. However, as the results of the current study will show, this claim may need to be modified because an implicit learning condition does not guarantee that learning will happen implicitly.

### Working memory

Working memory is a multicomponential construct that performs the dual function of information storage and processing. According to Baddeley (2015), it consists of a

central executive that coordinates the different components; a phonological store that encodes and rehearses verbal information; a visual-spatial store that stores information relating to images, space, or location; and an episodic buffer that integrates verbal and visual information and links with long-term memory. Among the components, the episodic buffer is the least researched because “we still know relatively little about the buffer” and there are “no agreed methods of measuring such capacity” (Baddeley, 2015, p. 26). Two major streams of research have emerged (Wen, 2015). One, led by Baddeley, focuses on the role of the phonological store in vocabulary learning (e.g., Baddeley, Gathercole, & Papagno, 1998). In the literature, this component of working memory, which only assumes the storage function, is called phonological short-term memory. It is measured through simple tasks such as nonword recall or digit span asking learners to memorize artificial words or meaningless strings of symbols. The other stream of research was initiated by Daneman and Carpenter (1980), who emphasize the importance of simultaneous information storage and processing. In their approach, working memory is measured via complex tasks such as listening span, reading span, or operation span, which tap both the storage and the processing components. Following Wen (2015), we will call this type of working memory “executive working memory.”

To what extent does working memory affect L2 learning? In a meta-analysis based on 748 correlation coefficients reported in 79 studies, Linck, Osthus, Koeth, and Bunting (2014) reported an overall weak correlation between all (both simple and complex) measures of working memory and second language proficiency,  $r = .25$ ; executive working memory was found to be more predictive of learning than phonological short-term memory,  $r = .27, .17$ , respectively; verbal measures demonstrated greater predictive power than nonverbal measures,  $r = .26, .18$ , respectively. However, the meta-analysis did not distinguish predictive studies that investigated the predictive power of working memory for ultimate learning success and interactional studies that examined the differential roles played by working memory under different learning conditions; nor did it examine whether working memory had differential effects on specific aspects of language learning such as L2 skills—reading, listening, writing, and speaking—and L2 knowledge—vocabulary and grammar. The following review will center on primary studies investigating the role of working memory in grammar learning, especially studies examining the interaction between treatment type and working memory.

Predictive research has shown that phonological short-term memory and executive working memory were both correlated with grammar learning (Engel de Abreu & Gathercole, 2012; Hummel, 2009), even after partialing out the influence of learners’ previous grammar knowledge (French & O’Brien, 2008). However, Serafini and Sanz (2016) reported that working memory was more relevant for initial grammar learning and as learners moved to more advanced levels, its effects weakened—a pattern similar to vocabulary learning (Cheung, 1996). In their study, second language (L2) Spanish learners at beginning, intermediate, and advanced levels were tested on 10 grammatical features at three time points during and after a semester of instruction. It was found that phonological short-term memory and executive working memory were only predictive of the grammar knowledge of the beginning and intermediate learners, but not those of the advanced learners. Serafini and



Sanz's findings echo Li's (2015a) meta-analytic finding about the stronger influence of language aptitude on initial L2 learning than learning at more advanced stages. However, except for Serafini and Sanz's study, level or stage of learning has not been investigated as an independent variable in primary studies. Therefore, there is a clear need for more research in this area.

While the predictive studies that did not examine treatment effects showed significant correlations between the predictor and criterion variables, experimental studies revealed whether the effects of working memory had to do with the characteristics of the instructional treatments (Li, 2014, 2015b). Two experimental studies where learners had no previous knowledge about the target language showed that both phonological short-term memory and executive working memory were predictive of the effects of instructional treatments involving discrete item-based learning (Kempe, Brooks, & Khrkhourin, 2010; Martin and Ellis, 2012). In both studies, the instruction was computerized and learning happened inductively by understanding input materials and receiving feedback containing the correct linguistic models. However, the above two studies only examined one learning condition. Studies investigating the differential roles of working memory in multiple learning conditions allow us to have a clearer picture about what instructional characteristics may contribute to the presence and absence of working memory effects. In this regard, the studies on corrective feedback—a form-focusing strategy investigated in this study—are revealing.

Seven studies have examined the mediating role of working memory in affecting the effects of recasts. Among the studies, four (Goo, 2012; Kim, Payant, & Pearson, 2015; Révész, 2012; Trofimovich et al., 2007) reported significant correlations between working memory and the effects of recasts while three failed to detect any significant effects (Li, 2013a, 2013b; Yilmaz, 2013). Therefore, these studies are unable to provide an unequivocal answer to the question of whether recasts, a feedback type favored by cognitive interactionists such as Long (2015), implicate working memory. Nevertheless, regardless of whether a significant correlation was found, all the studies resorted to “noticing” to interpret their findings. Where there was a significant effect, it was argued that working memory facilitated the noticing of recasts; in the absence of a significant effect, the argument was that working memory was not implicated because the implicit nature of the recasts meant that the learners did not recognize their corrective force.

However, we are unsure about whether noticing can serve as a convincing basis for the interpretations of the above findings because in all these studies recasts intensively targeted a single linguistic structure, which made the treatments very salient and the corrective intention easily noticed. One alternative explanation is that there were heavy processing demands imposed on learners in those studies where working memory played a significant role. Processing demand refers to the amount and complexity of information that the learner has to attend to in a task. For example, in Kim et al. (2015) and Trofimovich et al. (2007), learners were required to recall each recast they received while performing the treatment task. In Goo (2012), learners were given heavy doses of recasts in discrete item practice. In Révész (2012), learners were given 40 s to describe each of the 10 photos and received feedback on their errors while performing the task. In contrast, in Yilmaz's (2013) and Li's (2013a, 2013b) studies, there was no evidence of the tasks

imposing a heavy processing burden in the recast conditions, and therefore no significant effects were found for working memory.

Two studies investigated the role of working memory in the effects that metalinguistic feedback had on the learning of complex linguistic structures—the English that-trace filter in Goo (2012) and the Chinese perfective *-le* in Li (2013b). The researchers stated that the structures involved complicated linguistic projections and were nonsalient and opaque in form–meaning mapping. While Goo reported a null effect for working memory, Li found a negative effect. Li also found that the effects of metalinguistic feedback on the learning of the complicated linguistic structure were positively correlated with language analytic ability. The results suggest that working memory has either no or a harmful effect but language analytic ability has a positive effect on the processing of complicated metalinguistic information provided in online corrective feedback. Li interpreted the negative effect of working memory as suggesting that (a) learners with greater working memory capacities but weaker analytic abilities are unable to extract linguistic regularities and (b) learners with stronger memory abilities tend to store linguistic stimuli as unanalyzed chunks without conducting deep cognitive processing of the structure of the received input.

Note that in the above studies, no pretask grammar instruction was provided. What evidence is there for the relationship between working memory and the effects of pretask grammar instruction—a focus of this study? Sanz, Lin, Lao, Stafford, and Bowden (2016) is the only study that has investigated this topic. The study consists of two experiments. In Experiment 1, learners received a 70-min computer-administered treatment on L2 Latin learning composed of (a) vocabulary presentation and testing, (b) a grammar lesson, and (c) item-based comprehension practice where learners received metalinguistic feedback on their errors. No significant correlations were found between working memory and the learners' gain scores. Experiment 2 followed the same procedure except that learners did not receive the pre-practice grammar lesson. Significant correlations were found for both immediate and delayed gains after the treatment. The researchers interpreted the results as suggesting that (a) explicit grammar instruction cancelled the effect of working memory, and (b) working memory is “the key factor associated with success in less explicit language learning conditions” (p. 688).

In sum, working memory appears to be a weak, albeit significant, predictor of L2 learning, which contrasts with language analytic ability, a strong predictor of learning outcomes. However, with the caveat that more research is needed, there is evidence showing that as learners advance to higher stages of learning, the influence of working memory on learning outcomes may decline. The treatment studies demonstrate that similar to language analytic ability, the function of working memory varies under different learning conditions. For example, providing grammar instruction before practice activities may alleviate learners' processing burden and therefore cancel the effects of working memory. The feedback studies showed mixed findings, and most feedback researchers resort to the noticing function of working memory when interpreting their results. However, we offer an alternative interpretation, namely, that it is the heavy processing load imposed on the learners that led to the significant effects of working memory. We will revisit and elaborate on this argument in the Discussion section.



## Measurement of L2 Knowledge

The findings of experimental studies may be affected by the way treatment effects are measured. For example, Li's meta-analysis (2010) showed that corrective feedback had larger effects on free production measures, which might be partly because of the compatibility between the instruction type (free oral production tasks) and outcome measures. In a similar vein, Norris and Ortega's (2000) research synthesis shows that L2 instruction demonstrated superior effects on measures of explicit knowledge than on measures of implicit knowledge, which is likely due to the fact that in most studies, treatment effects are measured using tests favoring explicit knowledge. According to Ellis (2005), explicit knowledge is conscious, metalinguistic, and available for use in controlled processing, while implicit knowledge is unconscious, tacit, and accessible for spontaneous use. Ellis validated grammaticality judgment as a measure of explicit knowledge and elicited imitation as a measure of implicit knowledge. Ellis's model of explicit and implicit knowledge has been further confirmed in subsequent studies (Bowles, 2011; Kim & Nam, 2016; Zhang, 2015). However, Suzuki and DeKeyser (2016) challenged Ellis's model by providing evidence that elicited imitation is a measure of automatized explicit knowledge and that word monitoring is a better measure of implicit knowledge. One methodological disparity between Suzuki and DeKeyser's study and the studies following Ellis's model is that in Suzuki and DeKeyser's study, during the elicited imitation test the learners were told to correct grammar errors when repeating the sentences they were presented with, which may have increased the learners' chances of accessing their explicit knowledge.

The present study used two tests to measure treatment effects: grammaticality judgment and elicited imitation, which were designed as measures of explicit and implicit knowledge, respectively. The rationale for using the two tests is that cognitive ability may impact the learning of different knowledge types in different ways. For example, Révész (2012) reported that phonological short-term memory was only correlated with learners' scores on an oral test but not on a written test, while the opposite was true for executive working memory. Révész pointed out that this may suggest that phonological short-term memory is important for the acquisition of implicit knowledge while executive working memory is facilitative of explicit knowledge. However, most previous studies did not distinguish between the two types of knowledge. In the study reported here, we addressed this issue.

## The present study

Previous interactional aptitude studies have generated valuable findings regarding the interface between cognitive ability and treatment type. They demonstrate the fruitfulness of an interactional approach to cognitive aptitudes in revealing the mechanisms underlying second language acquisition and the joint contributions of learner-external factors (instruction) and learner-internal factors (cognitive factors) to L2 development. However, the studies also point to a number of issues that need to be addressed in further research. First, previous studies show that the two cognitive variables in question may interact with different instructional treatments in different ways. However, these studies investigated either only a limited

number of treatment types—mostly less than two—or only one cognitive variable. In addition, most studies only examined one type of form-focused instruction, not a combination of different options. Second, in most studies treatment effects were measured via tests that tapped into either explicit or implicit knowledge, and even when multiple tests were included, they did not clearly distinguish between these two types of knowledge. It remains to be seen whether the effects of language analytic ability and working memory are reflected differently in tests of the two types of knowledge. Third, the tests used to measure the two cognitive variables are diverse, making it difficult to compare the findings of different studies. Working memory has been measured via complex tasks such as listening or reading span tests gauging both the storage and the processing components, but in most studies only the recall (or storage) component was used as a predictor. However, it is unclear whether the recall component represented both information storage and information processing. Fourth, most of the studies were conducted in laboratory settings, raising questions about the generalizability of their findings to the classroom. Fifth, in most studies learners' previous knowledge was not included as a covariate, making it difficult to examine the unique contributions of these cognitive variables to new learning.

This study aims to address the above issues and gaps. It examines whether language analytic ability and working memory have differential associations with the learning outcomes of five types of instructional treatment distinguished by whether and when learners' attention is directed to the linguistic target (English past passive). The five conditions are as follows:

1. Explicit Instruction + Task: the learners received pretask grammar instruction before performing two communicative tasks.
2. Interactional Feedback: the learners received corrective feedback during task performance.
3. Explicit Instruction + Interactional Feedback: the learners received both pretask instruction and within-task feedback.
4. Posttask Feedback: the learners received corrective feedback (similar to interactional feedback) after completing the communicative tasks.
5. Task Only: the learners only performed the two communicative tasks and did not receive any form-focused instruction.

Accordingly, the following two research questions were formulated:

- RQ1: What is the relationship between language analytic ability and the immediate and delayed effects of the five instructional treatments?
- RQ2: What is the relationship between working memory and the immediate and delayed effects of the five instructional treatments?

To ensure the internal and external validity of the research, the study: (a) employed valid measures such as the PLAB 4 as the test of analytic ability, a grammaticality judgment test (GJT) as a test of explicit knowledge, and an elicited imitation test (EIT) as a test of implicit knowledge; (b) used composite scores combining the storage and processing components of working memory rather than just

recall scores; (c) included pretest scores as a predictor variable to tease out the influence of learners' previous knowledge on treatment effects; (d) opted for hierarchical regression analysis rather than simple correlations to determine the joint and independent contributions of the predictor variables; and (e) aimed for ecological validity by conducting the study in a classroom setting.

## Method

### Participants

The participants were 150 eighth-grade English as a foreign language learners at a Chinese middle school. Their average age was 14.05 years ( $SD = 0.54$ ), ranging from 13 to 15. These learners reported having studied English for an average of 6.24 years ( $SD = 1.38$ ). They were beginning English learners in terms of overall proficiency based on the requirements for English proficiency formulated by the Chinese Ministry of Education, and they also had limited previous knowledge about the target structure. They were recruited from 5 intact classes out of a total of 18 classes at the eighth grade. Each class comprised 60 students, and 30 were randomly selected as participants of the study. According to the school curriculum, the learners attended seven 45-min English lessons on a weekly basis, which featured explicit grammar instruction, rote learning, and mechanical practice. The learners were randomly assigned to five learning conditions, which were labeled Explicit Instruction + Task, Interactional Feedback, Explicit Instruction + Interactional Feedback, Posttask Feedback, and Task Only. One-way analyses of variance (ANOVAs) showed no significant differences between the five groups in their mid-term exam scores (their final exam scores were not yet available at the time of this study),  $F(4, 147) = 0.15$ ,  $p = .96$ , in their pretest scores on the GJT,  $F(4, 141) = 0.74$ ,  $p = .57$ , or in their pretest scores on the EIT,  $F(4, 137) = 1.05$ ,  $p = .38$ , suggesting that the groups were equivalent in their general proficiency and their previous knowledge of the target structure.

### Target structure

The target structure is the English past passive. We chose the passive voice as the target structure primarily because it was a new structure that had not been taught prior to this study. Targeting a new linguistic structure minimizes the effects of extraneous variables such as previous instruction and learning experience on the relationship between treatment effects and cognitive differences, thus increasing the robustness of the findings. According to Yalçın and Spada (2016), the passive voice is a difficult structure because (a) it involves multiple components (the auxiliary "be," the past participle of the verb, and the optional prepositional phrase "by + agent"), (b) it is infrequent in the input that learners receive in the classroom, and (c) it is a relatedly late-acquired structure. Yalçın and Spada (2016) reported that the learning of the English passive voice required language analytic ability, and in light of their finding, we expected a similar finding for our study. We further examined whether different instructional treatments manipulated on the basis of the timing of

form-focused instruction implicated the two cognitive factors in different ways in the learning of this linguistic feature.

### ***Instructional treatment***

Each of the five treatment groups performed two dictogloss tasks, but they differed in terms of whether and when they received form-focused instruction. The Explicit Instruction + Task group were given a short grammar lesson on the linguistic target (the English past passive) before performing the two tasks. The grammar lesson was composed of rule explanation followed by practice activities aiming to consolidate the newly learned grammar knowledge. The Interactional Feedback group did not receive pretask grammar instruction, but they received corrective feedback on their erroneous production of the target structure during task performance. The feedback consisted of a prompt encouraging the learner to self-correct, followed by a recast in the absence of self-correction. Thus under this treatment condition, form-focused instruction was embedded in meaning-primary tasks. The Explicit Instruction + Interactional Feedback group received both pretask instruction and within-task interactional feedback. In this treatment, form-focused instruction occurred both before and during task performance. The Posttask Feedback group received corrective feedback after completing the tasks, but they received no pretask instruction or within-task feedback. The procedure of the posttask feedback is similar to that of interactional feedback. Thus, in this learning condition, form-focused instruction was postponed until the posttask stage. While all the above treatment conditions involved some type of form-focused instruction, the Task Only group only performed the two communicative tasks without any form-focused instruction, so this condition was assumed to be purely meaning oriented.

### ***Dictogloss tasks***

A dictogloss task is one in which learners are asked to listen to a text presented by the teacher and then work in groups to reconstruct it. Dictogloss tasks integrate input and output activities and are especially useful for beginners and for teaching new linguistic structures. They can be implemented in different ways. In this study, each treatment group performed two dictogloss tasks involving the same procedures. The teacher started each task by going through a list of words that would appear in the narrative text. She then asked two brainstorming questions to arouse the learners' interest and activate their schematic knowledge about the topic. Following the initial warm-up, the teacher read the story aloud three times, first at normal speed to provide the learners with an initial idea, then at a slower pace with each sentence presented and annotated on PowerPoint slides to facilitate their understanding of the content and language, and finally at normal speed again to solidify their understanding. After listening to the story, the students were allowed 10 min to work in pairs and practice retelling it with a list of key words.<sup>1</sup> Each pair was also asked to add an ending to the story to increase within-group negotiation and to motivate the learners to listen to each other's narratives during the reporting stage. Upon completion of pair work, members of each pair were nominated to report the story to the rest of the class, with each member telling half of the story

before passing the speaker's role to his or her partner, who completed the whole narrative. While one group was telling the narrative, others were asked to predict how the speakers would end the story and compare the endings of each group. Following the reporting phase, the whole class was asked to vote for the best ending. On average, the dictogloss tasks in the five treatment conditions lasted about 38 min each.

Two narrative texts were used, one for each task. One text, which was composed by the authors, was about a car accident; the other, which was adapted from a story from *Reader's Digest*, reported an earthquake in Haiti. The lexical profiles of the two texts were compared using an online computer program (<http://www.lex tutor.ca/>), and they were found to be comparable in terms of lexical variety and density. The words in the two texts were all among the most frequent two thousand words, and any words not found in the learners' current and previous textbooks were pretaught and glossed. Two local teachers were consulted to ensure that the texts were appropriate for this level of learners in terms of length, difficulty, and topic. Fifteen exemplars of English past passives were embedded in each of the two texts.

#### *Explicit instruction*

The Explicit Instruction + Task and Explicit Instruction + Interactional Feedback groups both received a 10-min grammar lesson on the target structure before performing the oral narrative tasks. Following skill acquisition theory, the lesson included two components resembling the first two Ps of the PPP (presentation–practice–production) approach: rule explanation that aimed to provide learners with declarative knowledge, followed by controlled practice to deepen learners' understanding. For the rule explanation, the teacher asked students to identify the agent and the receiver of a passive sentence. She then explained and exemplified how the passive construction was formed and in what situations it was appropriate to use the structure. For the controlled practice, the teacher showed some sentences (e.g., "The bird was keep in a small cage") on PowerPoint, asking the whole class to judge whether the sentences were grammatical and correct the mistake if a sentence was ungrammatical. Students volunteered answers and received oral feedback on their answers to each item.

#### *Corrective feedback*

Corrective feedback was provided in three treatment conditions: Interactional Feedback, Explicit Instruction + Interactional Feedback, and Posttask Feedback.<sup>2</sup> The feedback, which was a variant of "corrective recasts" (Doughty & Varela, 1998), consisted of two moves: teacher repetition of error followed by reformulation of the wrong utterance if a student failed to self-correct. With regard to interactional feedback, when a learner produced a wrong passive sentence during the reporting stage, the teacher repeated the sentence with a rising intonation and with prosodic emphasis on the error to alert the learner to the presence of the error and to elicit self-correction. No further assistance was provided if the error was self-corrected. If the error remained, the teacher reformulated the sentence, replacing the error with the correct form without changing the meaning. Posttask feedback was provided

after all students had completed both tasks. The feedback was provided on errors logged by the teacher and another researcher while the learners were telling the narratives to the rest of the class. The teacher started each corrective episode by quoting a wrong sentence with prosodic emphasis on the error, followed by a prompt to elicit self-correction (e.g., “John, you said, ‘Three people were kill.’ Can you say it again?”). The teacher moved on to the next error if the learner was able to self-correct; otherwise, she provided a recast. The feedback session lasted about 15 min.

### ***Measures of the independent and dependent variables***

The independent or predictor variables were language analytic ability and working memory. Language analytic ability was measured using the language analysis subtest of the PLAB, which was validated on 6,000 foreign language learners and has demonstrated strong predictive validity (Pimsleur *et al.*, 2004). Unlike the well-known MLAT, which targets college students, the PLAB is intended for use in Grades 7–12 and is therefore ideal for this study. The test started with a brief study phase where learners were initially provided with two words and one sentence in an artificial language and their English translations. They were then asked to translate an English sentence into the artificial language based on their understanding of the principles and rules represented in the given materials. Afterward, they were given the correct answer and an explanation of the answer. Learners were then given more vocabulary and example sentences before moving on to the test phase. The test phase included 15 items, each with a simple English sentence and four possible translations in the artificial language, from which learners were asked to choose the correct translation. The learners were allowed 15 min to complete the test. The internal reliability of the test, represented by Cronbach’s  $\alpha$ , was 0.78.

It is necessary to clarify that the language analysis test is essentially language neutral, and that the learners’ test scores were not affected by their proficiency in English. This argument is supported by the following evidence. First, all test instructions were translated into the learners’ first language. Second, the learning material only consists of four content words: “father,” “see,” “horse,” and “carry.” The test was piloted with 12 learners from the same cohort, who reported no difficulty with the four vocabulary items. Third, probably the strongest evidence is the fact that the learners’ mean score was higher than their native speaker counterparts reported in Pimsleur *et al.*’s manual for the PLAB (2014; see the Results section for further details).

Working memory was measured by means of an operation span test, which has been extensively used in psychological and second language acquisition research and has strong predictive validity (Goo, 2012; Suzuki & DeKeyser, 2016; Yilmaz, 2013). During the test, the learners responded to sets of math–letter combinations, each consisting of a simple math equation followed by an English letter (e.g.,  $[3 \times 4] - 2 = 6?$  F). For each item, the learner was asked to judge whether the math equation was correct and respond as quickly as possible while trying to remember the letter. After completing the whole set, the learner was asked to recall the letter for each item. The test had 75 items divided into 15 sets, with the number of items in each set ranging from three to seven. Each set or span size (3, 4, 5, 6, and 7) appeared



three times in the test. Each learner's working memory performance was represented by a composite score of all three components of the test stimuli: veracity judgment, reaction time, and letter recall. The composite score was calculated by averaging the *z* scores for the three components; *z* scores were calculated for each learner and for the whole sample. Veracity judgment and reaction time were included because they represent the processing dimension of working memory and because previous research showed a trade-off between the processing and storage (recall) dimensions (Leeser, 2007; Waters & Caplan, 1996). It took the learners around 10 min to complete the test. The internal reliability for the three components was 0.91, 0.72, and 0.94 for letter recall, veracity judgment, and reaction time, respectively. All instructions were given in the learners' native language, and as the stimuli consisted of math equations and single letters, the test is language neutral. Therefore, the learners' test performance was not influenced by their L2 proficiency.

The dependent variables were the learning outcomes measured by an untimed GJT and EIT, which were designed to assess learners' explicit and implicit knowledge, respectively. The GJT included 40 items, out of which 30 were target items and 10 were distractors. For each item, the learner was asked to judge whether it was grammatical and correct the error if it was ungrammatical. Because previous research shows that ungrammatical items are more likely to tap explicit knowledge (Gutiérrez, 2013), all the target items were ungrammatical. One point was given if an ungrammatical sentence was judged as ungrammatical, and the error with passive use was corrected. If the error was not corrected, the item received a zero regardless of whether the judgment was correct. The internal reliability of the GJT was 0.91. The test lasted approximately 15 min.

In the EIT, learners were asked to listen to 40 statements (30 target items and 10 distractors) presented via DMDX, a computer program widely used in psycholinguistic research to present aural or visual stimuli. For each item, the learners were asked to first judge whether a statement was true of his or her personal life (e.g., "I was given a nice present on my birthday") and then repeat the sentence in correct English. The learners were not informed of the nature of the stimuli, namely, that some sentences were grammatical and some were ungrammatical. To restrict learners' access to explicit knowledge, a time limit was imposed on each item. The duration for each item was validated by asking 26 nonparticipant students to take the test and setting the time allowed for each item by calculating the average time it took the learners to complete each item. The average duration allowed for the test items was 6.2 s.<sup>3</sup> Because previous research showed no difference in learners' accuracy for grammatical and ungrammatical sentences in an EIT (Erlam, 2006), half of the items were grammatical and half ungrammatical. The ungrammatical items in both the GJT and the EIT were created based on three types of errors that emerged in the task performances of 24 learners in a pilot study. The error types were (a) bare verb (e.g., "My friend injure last week"); (b) no past participle (e.g., "My friend was injure last week"); and (c) no be (e.g., "My friend injured last week"). One point was given for each correct passive use, and responses containing incorrect use of the passive construction received a zero. Errors that were unrelated to the target structure were not scored. The internal validity of the EIT was 0.68. The test took each student 8–10 min to complete.

**Table 1.** Descriptive statistics for language analytic ability and working memory

Group	Language analytic ability			Working memory		
	<i>n</i> <sup>a</sup>	<i>M</i> <sup>b</sup>	<i>SD</i>	<i>n</i>	<i>M</i> <sup>c</sup>	<i>SD</i>
All groups	149	8.15	3.64	144	.03	.67
Explicit Instruction + Task	30	8.63	3.53	29	-.03	.69
Interactional Feedback	29	7.83	3.55	30	.12	.68
Explicit Instruction + Interactional Feedback	30	8.17	3.79	30	.07	.73
Posttask Feedback	30	8.63	3.56	28	.07	.53
Task Only	30	7.50	3.86	27	-.11	.73

Note: <sup>a</sup>The number of participants ranges from 27 to 30 per group and test after outliers were removed. <sup>b</sup>The maximum score is 15. <sup>c</sup>This is the composite of the standardized scores of the three elements of the working memory test: veracity judgment, reaction time, and letter recall.

### Procedure

This large-scale study is a cross-sectional study where learners received a one-shot treatment and took tests of treatment effects and cognitive abilities. The study spanned 3 weeks and was conducted in three sessions outside the learners' regular class hours in their regular classroom settings.<sup>4</sup> Session 1 lasted approximately 1 hr, and in this session, the learners took the working memory test, the language analytic ability test, the GJT pretest, and the EIT pretest. In Session 2, the learners participated in a 2-hr treatment during which they performed two dictogloss tasks according to the form-focused treatment they had been allocated, followed by the immediate GJT and EIT posttests. All instructional treatments were implemented by the same teacher, who is one of the authors of this article. Two weeks after the treatment, in Session 3, all learners took the delayed posttest. A delayed posttest was administered because previous research showed that the effects of cognitive variables may depend on the timing of the assessment of treatment effects (Li, 2013b).

## Results

### Descriptive statistics

#### *The predictor variables*

The learners' performances on the two aptitude tests appear in Table 1, including the mean scores and standard deviations for the whole sample and for the different participant groups. As a whole cohort, the learners scored an average of 8.15 out of 15 on the test of language analytic ability, which is higher than the mean score of their US counterparts (i.e., eighth graders;  $M = 6.8$ ,  $N = 979$ ) as reported in the PLAB Manual (Pimsleur *et al.*, 2004). A one-way ANOVA did not detect significant differences between the different groups in their scores on this test,  $F(4, 148) = 0.55$ ,  $p = .70$ . The learners' working memory scores were the average  $z$  scores of all three components—veracity judgment, reaction time, and letter recall.

**Table 2.** Descriptive statistics for grammaticality judgment

Group	Pretest			Posttest 1			Posttest 2		
	<i>n</i> <sup>a</sup>	<i>M</i> <sup>b</sup>	<i>SD</i>	<i>n</i>	<i>M</i>	<i>SD</i>	<i>n</i>	<i>M</i>	<i>SD</i>
Explicit Instruction + Task	29	2.03	2.97	30	10.10	8.93	30	10.57	9.01
Interactional Feedback	28	1.89	3.30	30	9.43	8.19	29	6.90	8.20
Explicit Instruction + Interactional Feedback	29	2.07	2.89	29	14.97	9.63	30	12.43	11.03
Posttest Feedback	29	1.31	1.79	30	8.70	9.45	28	3.79	4.10
Task Only	28	1.29	1.80	29	4.62	5.69	29	5.03	6.79

Note: <sup>a</sup>The number of participants ranges from 28 to 30 per group and test after outliers were removed. <sup>b</sup>The maximum is 30.

One-way ANOVA showed no significant differences between the five groups in their working memory scores,  $F(4, 143) = 0.52, p = .72$ .

#### *The criterion variable*

The criterion or dependent variable is the immediate and delayed GJT and EIT scores after the instructional treatments. Following studies with a similar design (Erlam, 2005; Suzuki & DeKeyser, 2016), only descriptive statistics are reported here, and no inferential statistical analyses were conducted, because the primary focus of the study is on the associations between cognitive aptitudes and learning gains under different learning conditions rather than differences in learning outcomes. Table 2 reports the learners' GJT scores. The pretest scores ranged from 1.29 to 2.07 out of 30, suggesting that the learners had very limited knowledge about the target structure at the outset of the study. At the time of the immediate posttest, all the experimental groups scored higher in comparison with their pretest performances. However, the group receiving both pretask instruction and within-task feedback improved the most, and the Task Only group the least. At the time of the delayed posttest, the two groups with explicit instruction scored higher than the other groups. The mean score of the Task Only group increased from the immediate posttest while the two feedback groups showed substantial decreases.

Table 3 displays the learners' EIT scores before and after receiving the treatments. Similar to the GJT, the EIT showed that the learners had limited previous knowledge prior to the study. In addition, overall the five groups demonstrated limited gains from the pretest to the two posttests; the gains were mostly below three points. The Explicit Instruction + Interactional Feedback group had higher posttest scores than the other groups on both posttests.

#### ***The associations between cognitive aptitudes and treatment conditions***

The associations between the two cognitive variables and the effects of different instructional treatments were investigated through multiple regression analysis, which shows the joint and unique contributions of each predictor to the outcome variable. In each analysis, the predictor/independent variables were language

**Table 3.** Descriptive statistics for elicited imitation

Group	Pretest			Posttest 1			Posttest 2		
	<i>n</i> <sup>a</sup>	<i>M</i> <sup>b</sup>	<i>SD</i>	<i>n</i>	<i>M</i>	<i>SD</i>	<i>n</i>	<i>M</i>	<i>SD</i>
Explicit Instruction + Task	27	1.70	2.13	29	2.69	2.93	29	3.57	3.16
Interactional Feedback	27	1.78	1.87	28	2.86	2.26	29	4.10	3.45
Explicit Instruction + Interactional Feedback	29	1.86	1.98	29	4.14	3.62	29	5.07	4.23
Delayed Feedback	26	1.19	1.49	28	2.96	2.67	29	3.31	3.12
Task Only	29	2.28	2.28	29	3.03	2.43	30	4.00	2.85

Note: <sup>a</sup>The number of participants ranges from 28 to 30 per group and test after outliers were removed. <sup>b</sup>The maximum is 30.

analytic ability, working memory, and pretest scores, and the dependent/outcome variable was learners' immediate or delayed posttest scores on the GJT or the EIT. Pretest scores were included as a predictor variable because previous research (Yalçın & Spada, 2016) showed that pretest scores accounted for a significant portion of the variance of posttest scores.<sup>5</sup> Posttest scores rather than gain scores were used as dependent variables because of the possible limitations of gain scores in undermining the robustness of the results (see Discussion). To examine whether the two cognitive variables were significant predictors of the outcome variable after controlling the effects of pretest scores, hierarchical regression analyses were conducted with pretest scores always entered at the first step and the two cognitive variables added at subsequent steps. To examine whether the two cognitive variables made unique contributions to the outcome variable, each was added at a different step, and the analysis was repeated with the two variables entered in a reverse order. The assumptions for multiple regression analysis were checked, and no violation of the assumptions was found (Field, 2009). Specifically, the Durbin–Watson tests indicated no violation of the “independence of errors” assumption, with all the test statistics falling between 1 and 3. The variance inflation factor was examined for each regression model as a diagnostic test of multicollinearity, but no cause for concern was detected (no variance inflation factor value was above 10). Furthermore, to ensure the robustness of the results, outliers were consistently removed. That is, any data point 3 *SD* units above or below the mean score was excluded from the analysis. Finally, the reported results for each analysis include (a) the predictors remaining in the final, best fitting model, (b) the standardized coefficient  $\beta$  for each predictor, which indexes the change in the outcome in *SD* units if the predictor increases by 1 *SD* unit, and (c) the adjusted  $R^2$  value, which represents the amount of variance accounted for by the predictors.

#### *Best fitting prediction models for the GJT scores*

Table 4 presents the best prediction model for each group's GJT scores on the immediate and delayed posttests. Starting from the Explicit Instruction + Task condition, the posttest scores under this learning condition were only predicted by the

**Table 4.** Significant predictors for grammaticality judgment scores

Group	Posttest 1			Posttest 2		
	Predictors	$\beta$	$R^2$	Predictors	$\beta$	$R^2$
Explicit Instruction + Task	Pretest	.39	.12	Pretest	.45	.17
Interactional Feedback	WM	.41	.60	WM	.56	.29
	Pretest	.52				
Explicit Instruction + Interactional Feedback	WM	.44	.42	WM	.44	.52
	Pretest	.38		Pretest	.45	
Posttask Feedback	LAA	.42	.14	Pretest	.53	.25
Task Only	LAA	.62	.35	LAA	.50	.22

Note: Pretest, pretest scores. WM, working memory. LAA, language analytic ability,  $\beta$ , standardized coefficient.  $R^2$ , adjusted  $R^2$  (amount of variance explained).

learners' pretest scores with neither of the two cognitive variables entering the analysis. For the Interactional Feedback group, working memory and pretest scores were significantly associated with the immediate effects and working memory was the sole significant predictor for the delayed effects. Similar to Interactional Feedback, the effects of Explicit Instruction + Interactional Feedback also showed significant relationships with working memory on both posttests. In addition, pretest scores explained a unique portion of the variance of both posttest scores. With regard to Posttask Feedback, the immediate effects were predicted by language analytic ability and the delayed effects by pretest scores. Finally, both the immediate and delayed posttest scores of the Task Only group were significantly predicted by language analytic ability.

To summarize the GJT results, the following patterns emerged from the data:

1. Language analytic ability was significantly predictive of the effects of Task Only and Posttask Feedback but not of the effects of other instructional treatments.
2. Working memory was a significant predictor for the two groups receiving interactional feedback but not for the other groups.
3. Neither language analytic ability nor working memory explained a significant amount of variance for the Explicit Instruction + Task group.
4. Pretest scores were a significant predictor of 6 out of the 10 prediction models.

#### *Best fitting prediction models for the EIT scores*

Table 5 shows that pretest scores were the most predominant predictor variable. It appeared in 9 out of the 10 prediction models and as a single significant predictor in 8. When it was a co-predictor, it showed greater predictive power than other predictors (i.e., it had larger  $\beta$  values). In addition, most of the  $\beta$  values for pretest scores were larger in comparison with the GJT results, suggesting its greater influence on the EIT results. Language analytic ability was not predictive of any EIT scores, but

**Table 5.** Significant predictors for elicited imitation scores

Group	Posttest 1			Posttest 2		
	Predictors	$\beta^c$	$R^2$	Predictors	$\beta$	$R^2$
Explicit Instruction + Task	Pretest	.65	.40	Pretest	.79	.61
Interactional Feedback	Pretest	.66	.41	WM	.37	.48
				Pretest	.47	
Explicit Instruction + Interactional Feedback	WM	.60	.33	WM	.38	.49
				Pretest	.43	
Posttask Feedback	Pretest	.76	.56	Pretest	.74	.52
Task Only	Pretest	.68	.44	Pretest	.74	.52

Note: Pretest, pretest scores. WM, working memory. LAA, language analytic ability.  $\beta$ , standardized coefficient.  $R^2$ , adjusted  $R^2$  (amount of variance explained).

working memory was a significant predictor of the delayed effects of Interactional Feedback and of both the immediate and the delayed effects of Explicit Instruction + Interactional Feedback—a pattern similar to the GJT results.

In sum, the elicited imitation test showed the following findings:

1. Working memory was a significant predictor for the two groups receiving interactional feedback, a pattern similar to the GJT results.
2. Language analytic ability was not predictive of any group's posttest scores, which contrasts with its significant associations with Task Only and Posttask Feedback on the GJT.
3. Pretest scores were a consistent predictor of all groups' posttest scores, and their predictive power was stronger for the EIT scores than the GJT scores.

## Discussion

This study sought to examine whether language analytic ability and working memory were differentially associated with the effectiveness of five different form-focused instruction options: Explicit Instruction + Task, Interactional Feedback, Explicit Instruction + Interactional Feedback, Posttask Feedback, and Task Only. The following results were obtained. First, language analytic ability was significantly predictive of the effects of Task Only and Posttask Feedback, but the significant associations were only found for the GJT, not the EIT. Second, working memory was implicated under the two conditions where interactional feedback was provided (Interactional Feedback and Explicit Instruction + Interactional Feedback), and the significant effects were found for both the GJT and the EIT. Third, learners' prior knowledge was a strong and consistent predictor of all treatment types. These findings for the two cognitive variables are summarized in Table 6. Below we discuss the major findings for the three predictors—language analytic ability, working memory, and pretest scores—with reference to the



**Table 6.** Language analytic and working memory in the five treatment conditions

Groups	Language analytic ability	Working memory
Explicit Instruction + Task	NO	NO
Interactional Feedback	NO	YES <sup>a</sup>
Explicit Instruction + Interactional Feedback	NO	YES <sup>b</sup>
Posttask Feedback	YES <sup>c</sup>	NO
Task Only	YES <sup>d</sup>	NO

Note: <sup>a</sup>Both GJT and EIT. <sup>b</sup>Both GJT and EIT. <sup>c</sup>Only GJT. <sup>d</sup>Only GJT.

methods of this study and previous research. We end this section with a brief discussion of the theoretical implications of the findings.

### **Language analytic ability**

We start by focusing on the results for the GJT and then explain why the significant effects were not found for the EIT. To begin with, the finding that analytic ability was drawn upon by learners who only performed the two meaning-oriented communicative tasks was unexpected. Previous studies showed that analytic ability was more likely to be implicated in explicit than in implicit learning conditions (Li, 2015a, 2017b, 2018; Sheen, 2007; Yilmaz, 2013). Given that these learners did not receive explicit instruction or corrective feedback, Task Only was the least explicit of all treatment conditions and the least likely to implicate an ability involving conscious learning. However, although Task Only was intended to be a meaning primary, implicit learning condition, several features of the treatment may have led the learners to engage in conscious learning. First, the English passive voice is a salient linguistic structure given that its linguistic projection is distinct from that of the canonical active voice, it is meaning distinctive, and it is transparent in form–meaning mapping. As Reber (1989) argued, implicit learning is less likely to happen when the linguistic target is salient, in which case learners are likely to engage in rule search. The second feature concerns the learners’ background or learning experience. These learners had been taught through a traditional, grammar-based approach, which may have oriented them toward a form-focused, conscious learning style even though the tasks they were asked to perform were meaning oriented. This argument rests on Ellis’s (2003, p. 5) distinction between “task-as-workplan” and “task-as-process,” with the former representing a task designer’s perspective and the latter a participant’s perspective. In other words, tasks are designed to be meaning focused, but whether or not the participants perform a task in the way desired by the designer is an open question. The third feature pertains to the focused nature of the dictogloss tasks, where the narrative texts included a total of 30 exemplars of the English passive voice. The frequent exposure to the target structure during the tasks may have induced awareness of the objectives

of the treatment and primed the learners to actively process the relevant input and extrapolate regularities.

The lack of significant effects for the two groups receiving pretask grammar instruction can be explained by the prior instruction neutralizing differences in the learners' analytic ability. This finding confirms Snow's (1991) argument that structured, external assistance favors low-aptitude learners. Another likely contributing factor is the production task following the explicit instruction, which Erlam (2005) argued may weaken the predictive force of analytic ability. One interesting explanation Erlam offered is that providing explicit instruction before production tasks is what usually happens in traditional language classes, thus increasing learners' motivation while minimizing the impact of aptitude. Unfortunately, neither her study nor this study included a measure of motivation. Despite the plausibility of the above explanations, the finding of this study differs from Yalçın and Spada's (2016), which reported significant links between analytic ability and the learning of the English passive voice resulting from a form-focused treatment seemingly similar to the two treatment conditions involving explicit instruction in this study. A closer inspection of the example instruction materials provided by Yalçın and Spada, however, indicates that the disparity between the findings of the two studies may have been due to the way the explicit instruction was implemented. In our study, the explicit instruction started with rule explanation followed by controlled practice consolidating the declarative knowledge—a deductive approach. In their study, the instruction started by providing learners with some input material, and then the teacher guided learners to discover the rule—an inductive approach (Cerezo, Caras, & Leow, 2016). As discussed in previous sections, inductive instruction is more likely to require analytic ability than deductive instruction.

While pretask grammar instruction likely wiped out the effects of analytic ability, we need to offer an explanation for why analytic ability was unrelated to the learning gains of the group that only received interactional feedback. Recall that the feedback provided in this study consisted of a prompt alerting the learners to the existence of an error and eliciting self-correction, followed by a recast providing the correct form if the learner failed to self-correct. The corrective intention was transparent, and therefore the feedback can be considered to be an explicit type. Li (2013b), Sheen (2007), and Yilmaz (2013) all reported that analytic ability was correlated with the effects of explicit feedback. However, in these studies the feedback types were essentially input providing (i.e., they provided the correct form). Therefore, the learners in these studies primarily engaged in input processing, and there was no demand for "pushed output" (Swain, 2005). The feedback in this study, however, included both input-providing and output-prompting moves, which required learners to both encode new information from the input and retrieve recently learned information. The heavy processing demands imposed by the prompting moves, which demanded immediate output and the instant application of the learned linguistic knowledge, may have made it difficult for learners to draw on their analytic ability and instead implicated working memory—a point to be revisited below.<sup>6</sup>

Next, analytic ability was predictive of the effects of Posttask Feedback, which involved a three-step procedure: (a) the teacher quoted an erroneous sentence produced by the learner during the narrative tasks and asked him or her to repeat it and

then self-correct; (b) the learner reproduced the sentence; and (c) the teacher reformulated the sentence if the error remained. As can be seen, although the feedback followed the same procedure as the interactional feedback provided during task performance, there were fundamental differences due to the inherent differences in the two instructional conditions. In the posttask feedback, there was no opportunity for immediate proceduralization or application of the learned linguistic knowledge in speech production, and the treatment was purely input based. In addition, errors were singled out and corrected in discrete sentences. In many respects, the treatment was similar to the structured input condition in Erlam (2005) and the rule search condition in Robinson (1997), both of which showed significant associations with analytic ability.

Finally, the fact that the effects of language analytic ability were only observed on the GJT but not the EIT suggests that this cognitive ability, which is important for conscious processing of linguistic material, facilitates the acquisition of explicit knowledge but not implicit knowledge. In this respect, the study is in line with Yalçın and Spada (2016), where analytic ability was only predictive of learners' GJT scores but not their performance on an oral test, which supposedly tapped implicit knowledge (Ellis, Loewen, & Erlam, 2006). It is also consistent with Li's studies (2013a, 2013b), which showed that language analytic ability was predictive of learners' GJT scores, not their EIT scores. It could also be argued that due to the short duration of the instructional treatments, the resulting declarative knowledge had yet to be proceduralized or automatized. The results in Tables 2 and 3 show that the learners' posttest scores were considerably higher than their pretest scores on the GJT, while the gains on the EIT were limited.

### **Working memory**

The most striking finding about working memory is that it was a predictor for the two groups who received interactional feedback but not for the other treatment groups. We argue that this is primarily due to the heavy processing burden imposed by the feedback on the learners' memory resources in real-time speech performance. As noted above, the feedback consisted of a prompt and a recast, which required the learner to retrieve information from long-term memory, compare the received input with his or her own output, encode new information in long-term memory, apply the information in immediate production, and at the same time plan the content and language of subsequent utterances. In the other treatment conditions—Explicit Instruction + Task, Posttask Feedback, and Task Only—learners did not have to handle online, continuous feedback. Our study, then, corroborates Suzuki and DeKeyser's (2016) finding that working memory was implicated in massed instruction, which posed heavier processing pressure but not in spaced instruction where the treatment was spread out and the processing load was less onerous. Based on the results of our and other studies, we argue that executive working memory, which involves the ability to manipulate and store information simultaneously, is needed in tasks that involve online or real-time information processing such as the focused communicative tasks. This argument partially aligns with Wen's (2015) recommendations for investigating (a) the role of executive working memory in mediating the processes of learning in production and comprehension tasks and (b) the role of phonological short-term

memory (only information storage) in affecting the final product of learning. While we are unsure about the matching between the two types of working memory and different aspects of learning, we agree with the need to investigate the role of working memory in instructional treatments based on clear theoretical account of the mechanism through which it affects learning.

One important point in the literature review is that previous researchers have resorted to the notion of noticing when interpreting the presence or absence of the effects of working memory. For example, in an attempt to explain why working memory was predictive of the effects of recasts but not those of metalinguistic feedback, Goo speculated that it was because the noticing of recasts “necessitates a domain-general, attention-control mechanism [the central executive]” (p. 466) in working memory whereas metalinguistic feedback did not require noticing because it was explicit. Kim *et al.* (2015) also noted “the benefits of high WM [working memory] on noticing of recasts” (p. 573). However, we would like to point out that in most feedback studies where a single structure receives intensive treatment and where the learners are likely to be oriented toward explicit learning, implicit instruction may not be perceived as implicit because the instructional objectives could be easily recognized. That being the case, the presence or absence of working memory effects may not be due to whether noticing is involved (Li, 2017b). In our study, given the explicit nature of the feedback, noticing is unlikely to account for the significant effects of working memory but processing load might.<sup>7</sup>

Unlike language analytic ability whose effects were only related to explicit knowledge, working memory, which assumes both storage and processing functions, has been found to be correlated with both explicit knowledge (Goo, 2012; Li, 2013b; Révész, 2012; Sheen, 2007) and implicit knowledge (this study; Li, 2013b; Kim *et al.*, 2015). This finding can be interpreted from two perspectives: testing and learning. From a testing perspective, learners with higher working memory are probably better at retrieving linguistic knowledge from their long-term memory, which is critical in spontaneous speech production. From a learning perspective, learners with more memory resources are probably better at internalizing linguistic input as unanalyzed chunks, which are readily available for automatic use. However, Révész (2012) distinguished phonological short-term memory that only taps the storage component and complex working memory that concerns the ability to simultaneously store and process information. Based on her findings, she argued that phonological short-term memory is critical for the development of implicit knowledge while complex working memory is important for explicit knowledge. Clearly, the relationship between working memory and L2 knowledge type needs to be further investigated.<sup>8</sup>

### **Prior knowledge**

We found that learners’ prior knowledge of the linguistic target, operationalized as pretest scores, was a consistent predictor for both GJT and EIT posttest scores. Including pretest scores as a variable made it possible to tease out the unique contribution of aptitude after accounting for the variance explained by learners’ previous knowledge. Language aptitude was conceived as learners’ initial readiness for learning. In the validation research for the MLAT—the most influential aptitude

test—Carroll (1990) excluded learners with prior language learning experience. In situations where it is difficult to find learners with zero knowledge of the target language or structure, as is the case with most aptitude studies, it is important to account for the influence of learners' prior knowledge on treatment effects. As this study has shown, even though the passive voice was a "new" structure, pretest scores were a strong and consistent predictor. Despite the importance of previous knowledge in affecting learning outcomes, to date only one aptitude study included pretest scores as a predictor variable (Yalçın & Spada, 2016).

Including pretest scores as a predictor or covariate may also enable us to avoid the controversy over whether gain scores are an appropriate outcome variable. A gain score represents the change or development of learners' L2 knowledge from Point A to Point B, and it is typically calculated by subtracting a pretest score from a posttest score. Gain scores favor learners with low pretest scores, they change the variance and distribution of raw scores, and in cases where a learner's posttest score is lower than his or her pretest score, a negative gain score is obtained. Gain scores have been criticized for having low reliability in comparison with raw scores (e.g., Cronbach & Furby, 1970). However, there has also been a counterargument that gain scores are equally reliable as raw scores (Zimmerman & Williams, 1982). It is beyond the scope of this article to resolve the controversy, but we would like to recommend using posttest scores as the criterion/outcome variable and pretest scores as a covariate, in order to avoid any potential pitfalls relating to the use of gain scores as a dependent variable. Most aptitude researchers use gain scores as the outcome variable probably to control for the influence of pretest scores, but pretest scores may still be a significant predictor of gain scores. In practical terms, this would mean that previous knowledge is associated with learning gains or that learners with more previous knowledge about the linguistic target learn more. Therefore, even when gain scores rather than posttest scores are used as an outcome variable, it is still necessary to investigate the impact of learners' prior knowledge.<sup>9</sup>

### **Theoretical implications**

Most second language acquisition theories are primarily concerned with how learning happens, and they do not make explicit claims about the role of learners' cognitive variation in affecting learning outcomes. Even if they do, they only make some overarching claims without elaborating whether and how cognitive aptitudes are implicated differently under different learning conditions. Skill acquisition theory holds that L2 learning involves the acquisition of declarative/explicit knowledge, which is subsequently proceduralized and ultimately automatized through practice. Based on this theory, learning is conscious and relies heavily on abilities for explicit learning such as analytic ability and working memory. However, our study shows that the mediating effects of aptitude disappear when deductive explicit instruction is provided. Cognitive interactionists (e.g., Long, 2015) advocate the use of online corrective feedback to enhance acquisition. However, recognition needs to be given to the cognitive burden that focus-on-form can incur. Krashen's (1981) input theory dismisses the importance of cognitive aptitudes in incidental learning. However, this study shows that learners apply their analytic ability and engage in rule learning when they perform communicative tasks

without receiving any form-focused instruction. As previously discussed, this may be due to the overall form-focused orientation of the instructional context. However, as Harley and Hart's study (1997) shows, the influence of analytic ability is evident even in immersion classes where learners receive consistent meaning-oriented instruction—an approach advocated by Krashen. Finally, our findings contribute to research on aptitude–treatment interaction by showing the differential effects of the two cognitive variables under different learning conditions (DeKeyser, 2012; Li, 2018). This study, therefore, affords further evidence for the benefits of an interactional approach to cognitive aptitudes in unraveling the mechanism of second language acquisition. We hope that the study serves as an impetus for further research along this line.

## Conclusion

This study constitutes the first attempt to examine the relationships between two important cognitive variables and the timing (and presence/absence) of form-focused instruction. The study shows that (a) the influence of analytic ability is evident when there is a lack of external assistance and when the instruction requires active processing of input materials, and (b) working memory is drawn upon in conditions where learners have a heavy cognitive load. The study was conducted using rigorous methods: it was based on a relatively large sample ( $N = 150$ ); the variables were carefully manipulated; the treatments were consistently implemented; and all variables were measured via validated tests. Therefore, the findings have high internal and external validity and contribute significantly to interactional aptitude research and second language acquisition research in general.

The study also has limitations. First, although previous studies have validated elicited imitation as a measure of implicit knowledge (Bowles, 2011; Elli, 2005; Kim & Nam, 2016; Zhang, 2015), the test, which requires learners to repeat aurally presented sentences in the L2 under time pressure, may have been too challenging for the young learners in this study. Future research may allow learners more time for the oral repetition part of each item or include alternative measures of implicit knowledge such as oral production tasks or word monitoring (Suzuki & DeKeyser, 2016). Second, an anonymous reviewer pointed out that although previous research shows that ungrammatical items are more likely to tap explicit knowledge, the decision not to include grammatical items as critical items in the GJT warrants further consideration. This concern needs to be addressed in future research. Third, although the study was carefully designed and strictly executed, it was conducted in a classroom setting where certain variables could not be easily controlled. For example, some students may not have been engaged in the instructional activities. One way to resolve this issue is to measure learners' motivation and include it as a covariate in the data analysis. Another way is to exclude those who did not engage with the instructional treatment. Fourth, the current study investigated the learning of a new structure by beginning L2 learners. However, previous research showed that cognitive ability may be less important for advanced learning (Sanz *et al.*, 2016). Future research may explore whether the findings of this study will be obtained at more advanced stages of learning.



## NOTES

1. Learner–learner interaction during the dictogloss tasks may have contributed to the effects of the instructional treatments, but this feature was the same in all treatment conditions, which differed only in whether and when learners received form-focused instruction. Therefore, an interesting feature to investigate the interaction between learners possibly did not have a direct impact on the relationships between the two cognitive factors and the effectiveness of the different treatments.
2. The number of feedback episodes provided was 50, 48, and 36 for Interactional Feedback, Explicit Instruction + Interactional Feedback, and Posttask Feedback, respectively.
3. One reviewer worried that the time limit allowed for the repetition may have been too short. However, a time limit must be imposed to restrict explicit knowledge. The time limit for each item was validated based on a pilot study; it was the time it took most learners to complete a particular item. In Kim and Nam's study (2016), the average time allowed for each item was 3.6 s, much shorter than the average time limit in this study (6.2 s).
4. One reviewer pointed out that given that the study was conducted outside the learners' regular class hours, it is not a pure classroom-based study, or it is better considered to be a lab-based study. Our view is that an accurate label for this study is perhaps "a quasi-classroom study." The study has some features of a classroom study because it was conducted with intact classes and the instructional activities mirror what happens in the classroom in the real world. However, the study is experimental and the treatments were strictly controlled, which makes it similar to a lab study in certain aspects.
5. One reviewer doubted whether it is reasonable to include pretest scores as a predictor given that there were significant differences between the participant groups. However, the regression analysis was conducted with each group's posttest scores, and lack of between-group differences does not preclude the influence of a particular group's previous knowledge on their posttest performance. Furthermore, even though the learners had limited previous knowledge about the target structure, their pretest scores varied. As the results showed, variation in the learners' pretest scores was a consistent predictor of their posttest scores.
6. One reviewer brought up the possibility that in the Interactional Feedback condition, while one student was receiving feedback during his/her oral narrative, other students might have engaged in analyzing the student's output and the feedback. This may have led to the use of the students' language analytic ability, and yet no effects for analytic ability were found. We speculate that this is because, as we explained in the article, the learners in this group received online feedback and had an opportunity to continuously apply the learned knowledge in immediate production. It is the requirement for production that cancelled the effects of analytic ability.
7. To clarify, we endorse the view that working memory is important for noticing the gap. However, the noticing function of working memory is irrelevant when the objectives of the instruction can be easily recognized, which is what happens in most treatment studies, including this one.
8. An anonymous reviewer inquired about the relationship between working memory and the short-term and long-term effects of instructional treatments. In the literature, some studies (e.g., Li, 2013a; Mackey et al., 2002) reported a significant effect of working memory on only the delayed effects of instructional treatments, while other studies including Li's (2017b) meta-analysis showed that working memory was predictive of the immediate, but not delayed, effects of L2 instruction. This study showed that working memory was a significant predictor of both the immediate and the delayed posttest scores of the two groups receiving online feedback. The disparities cannot be easily accounted for, and more research is in order to elucidate the interface between the sustainability of treatment effects and cognitive variables.
9. Previous knowledge may have a negative impact on gain scores in the case of ceiling effects, that is, learners may have too much knowledge about the linguistic target, thus leaving little room for further improvement.

## References

- Baddeley, A.** (2015). Working memory in second language learning. In M. Mota, and A. McNeill (Eds.), *Working memory in second language acquisition and processing* (pp. 17–28). Bristol: Multilingual Matters.
- Baddeley, A., Gathercole, S., & Papagno, C.** (1998). The phonological loop as a language learning device. *Psychological Review*, **105**, 158–173.

- Bowles, M.** (2011). Measuring implicit and explicit knowledge: What can heritage learners contribute? *Studies in Second Language Acquisition*, **33**, 247–271.
- Carroll, J.** (1990). Cognitive abilities in foreign language aptitude: Then and now. In T. Parry, and C. Stansfield (Eds.), *Language aptitude reconsidered* (pp. 11–29). Englewood Cliffs, NJ: Prentice Hall.
- Carroll, J., & Sapon, S.** (2002). *Manual for the MLAT*. Princeton, NJ: Second Language Testing.
- Cerezo, L., Caras, A., & Leow, R.** (2016). The effectiveness of guided induction versus deductive instruction on the development of complex Spanish *Gustar* structures. *Studies in Second Language Acquisition*, **38**, 265–291.
- Cheung, H.** (1996). Nonword span as a unique predictor of second-language vocabulary learning. *Developmental Psychology*, **32**, 867–873.
- Cronbach, J. L., & Furby, L.** (1970) How should we measure 'change'-or should we? *Psychological Bulletin*, **74**(1), 68–80.
- Daneman, M., & Carpenter, P.** (1980). Individual differences in working memory and reading. *Journal of Verbal Learning and Verbal Behaviour*, **19**, 450–466.
- DeKeyser, R.** (2012). Interactions between individual differences, treatments, and structures in SLA. *Language Learning*, **62**, 189–200.
- DeKeyser, R.** (2015). Skill acquisition theory. In B. VanPatten, and J. Williams (Eds.), *Theories in second language acquisition: An introduction*. New York: Routledge.
- Doughty, C., & Varela, E.** (1998). Communicative focus on form. In C. Doughty, and J. Williams (Eds.), *Focus on form in classroom second language acquisition* (pp. 114–138). New York: Cambridge University Press.
- Ellis, R.** (2003). *Task-based language learning and teaching*. Oxford: Oxford University Press.
- Ellis, R.** (2005). Measuring implicit and explicit knowledge of a second language: A psychometric study. *Studies in Second Language Acquisition*, **27**, 141–172.
- Ellis, R., Loewen, S., & Erlam, R.** (2006). Implicit and explicit corrective feedback and the acquisition of L2 grammar. *Studies in Second Language Acquisition*, **28**, 339–368.
- Engel de Abreu, P. M., & Gathercole, S. E.** (2012). Executive and phonological processes in second-language acquisition. *Journal of Educational Psychology*, **104**, 974–986.
- Erlam, R.** (2005). Language aptitude and its relationship to instructional effectiveness in second language acquisition. *Language Teaching Research*, **9**, 147–171.
- Erlam, R.** (2006). Elicited imitation as a measure of L2 implicit knowledge: An empirical validation study. *Applied Linguistics*, **27**, 464–491.
- Field, A.** (2009). *Discovering statistics using SPSS*. London: Sage.
- French, L. M., & O'Brien, I.** (2008). Phonological memory and children's second language grammar learning. *Applied Psycholinguistics*, **29**, 463–487.
- Goo, J.** (2012). Corrective feedback and working memory capacity in interaction-driven L2 learning. *Studies in Second Language Acquisition*, **34**, 445–474.
- Gutiérrez, X.** (2013). The construct validity of grammaticality judgment tests as measures of implicit and explicit knowledge. *Studies in Second Language Acquisition*, **35**, 423–449.
- Harley, B., & Hart, D.** (1997). Language aptitude and second language proficiency in classroom learners of different starting ages. *Studies in Second Language Acquisition*, **19**(3), 379–400.
- Hummel, K.** (2009). Aptitude, phonological memory, and second language proficiency in nonnovice adult learners. *Applied Psycholinguistics*, **30**, 225–249.
- Hwu, F., & Sun, S.** (2012). The aptitude-treatment interaction effects on the learning of grammar rules. *System*, **40**, 505–521.
- Hwu, F., Wei, P., & Sun, S.** (2014). Aptitude-treatment interaction effects on explicit rule learning: A latent growth curve analysis. *Language Teaching Research*, **18**, 294–319.
- Kempe, V., Brooks, P., & Kharkhurin, A.** (2010). Cognitive predictors of generalization of Russian grammatical gender categories. *Language Learning*, **60**, 127–153.
- Kim, J., & Nam, H.** Measures of implicit knowledge revisited: Processing modes, time pressure, and modality. *Studies in Second Language Acquisition*, **39**, 431–457.
- Kim, Y., Payant, C., & Pearson, P.** (2015). The intersection of task-based interaction, task complexity, and working memory. *Studies in Second Language Acquisition*, **37**, 549–581.

- Krashen, S.** (1981). Aptitude and attitude in relation to second language acquisition and learning. In K. Diller (Ed.), *Individual differences and universals in language learning aptitude* (pp. 155–175). Rowley, MA: Newbury House.
- Leeser, M.** (2007). Learner-based factors in L2 reading comprehension and processing grammatical form: Topic familiarity and working memory. *Language Learning*, *57*, 229–270.
- Li, S.** (2010). The effectiveness of corrective feedback in SLA: A meta-analysis. *Language Learning*, *60*, 309–365.
- Li, S.** (2013a). 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.** (2013b). The differential roles of language analytic ability and working memory 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 & language program administration: From theory to application* (pp. 32–52). Boston, MA: Cengage Learning.
- Li, S.** (2014). Recasts, working memory, and the choice of target structure. In Z. Han (Ed.), *Second language acquisition of Chinese: A series of empirical studies* (pp. 103–125). Bristol: Multilingual Matters.
- 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 working memory and language analytic ability in mediating the effects of recasts as a function of learner proficiency. In Z. Wen, M. Mota, and A. MacNeil (Eds.), *Working memory in second language acquisition and processing: Theory, research and commentary* (pp. 139–159). Bristol: Multilingual Matters.
- Li, S.** (2016). The construct validity of language aptitude. *Studies in Second Language Acquisition*, *38*, 801–842.
- Li, S.** (2017a). Cognitive differences and ISLA. In S. Loewen, and M. Sato (Eds.), *Handbook of instructed second language learning* (pp. 396–417). New York: Routledge.
- Li, S.** (2017b). The effects of cognitive aptitudes on the process and product of L2 interaction: A synthetic review. In L. Gurzynski (Ed.), *Expanding individual difference research in the interaction approach: Investigating learners, instructors and researchers* (pp. 41–70). Philadelphia: Benjamins.
- Li, S.** (2018). Language aptitude. In A. Burns, and J. Richards (Eds.), *The Cambridge guide to learning English as a second language* (pp. 63–72). Cambridge: Cambridge University Press.
- Linck, J., Osthus, P., Koeth, J. T., & Bunting, M.** (2014). Working memory and second language comprehension and production: A meta-analysis. *Psychonomic Bulletin & Review*, *21*, 861–883.
- Long, M.** (2015). *Second language acquisition and task-based language teaching*. Malden, MA: Wiley.
- Long, M.** (2016). In defence of tasks and TBLT: Nonissues and real issues. *Annual Review of Applied Linguistics*, *36*, 5–33.
- Mackey, A., Gass, S., & McDonough, K.** (2000). How do learners perceive international feedback? *Studies in Second Language Acquisition*, *22*, 471–497.
- Martin, K. I., & Ellis, N. C.** (2012). The roles of phonological short-term memory and working memory in L2 grammar and vocabulary learning. *Studies in Second Language Acquisition*, *34*, 179–413.
- Meara, P.** (2005). *LLAMA Language Aptitude Tests*. Swansea: Lognostics.
- Norris, J., & Ortega, L.** (2000). Effectiveness of L2 instruction: A research synthesis and quantitative meta-analysis. *Language Learning*, *50*, 417–528.
- Pimsleur, P., Reed, D., & Stansfield, C.** (2004). *Pimsleur Language Aptitude Battery (PLAB)*. Bethesda, MD: Second Language Testing.
- Plosky, L., & Oswald, F.** (2014). How big is “big”? Interpreting effect sizes in L2 research. *Language Learning*, *64*, 878–912.
- Reber, A.** (1989). Implicit learning and tacit knowledge. *Journal of Experimental Psychology: General*, *118*, 219–235.
- Révész, A.** (2012). Working memory and the observed effectiveness of recasts on different L2 outcome measures. *Language Learning*, *62*, 93–132.
- Robinson, P.** (1997). Individual differences and the fundamental similarity of implicit and explicit adult second language learning. *Language Learning*, *47*, 45–99.
- Sachs, R.** (2010). Individual differences and the effectiveness of visual feedback of reflexive binding in L2 Japanese (Unpublished doctoral dissertation, Georgetown University).

- Sanz, C., Lin, H., Lado, B., Stafford, C., & Bowden, H.** (2016). One size fits all? Learning conditions and working memory capacity in ab initio language development. *Applied Linguistics*, *37*, 669–692.
- Serafini, E., & Sanz, C.** (2016). Evidence for the decreasing impact of cognitive ability on second language development as proficiency increases. *Studies in Second Language Acquisition*, *38*, 607–646.
- Sheen, Y.** (2007). The effects 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). New York: Oxford University Press.
- Skehan, P.** (1996). A framework for the implementation of task-based instruction. *Applied Linguistics*, *17*, 38–62.
- Skehan, P.** (2012). Language aptitude. In S. Gass, and A. Mackey (Eds.), *Handbook of second language acquisition* (pp. 381–395). New York: Routledge.
- Snow, R. E.** (1991). Aptitude-treatment interaction as a framework for research on individual differences in psychotherapy. *Journal of Consulting and Clinical Psychology*, *59*, 205–216.
- Spada, N.** (2011). Beyond form-focused instruction: Reflections on past, present and future research. *Language Teaching*, *44*, 225–236.
- Suzuki, Y., & DeKeyser, R.** (2016). Exploratory research on second language practice distribution: An Aptitude  $\times$  Treatment interaction. *Applied Psycholinguistics*. Advance online publication. doi: 10.1017/S0142716416000084
- Swain, M.** (2005). The output hypothesis: Theory and research. In E. Hinkel (Ed.), *Handbook on research in second language teaching and learning* (pp. 471–484). Mahwah, NJ: Erlbaum.
- 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*. Oxford: Oxford University Press.
- Waters, G., & Caplan, D.** (1996). The measurement of verbal working memory capacity and its relation to reading comprehension. *Quarterly Journal of Experimental Psychology*, *49A*, 51–79.
- Wen, Z.** (2015). Working memory in second language acquisition and processing: The phonological/executive model. In Z. Wen, B. Mailce, and A. McNeill (Eds.), *Working memory in second language acquisition and processing*. Bristol: Multilingual Matters.
- Willis, D., & Willis, J.** (2007). *Doing task-based teaching*. Oxford: Oxford University Press.
- Yalçın, Ş., & Spada, N.** (2016). Language aptitude and grammatical difficulty: An EFL classroom-based study. *Studies in Second Language Acquisition*, *38*, 239–263.
- 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.
- Zhang, R.** (2015). Measuring university-level L2 learners' implicit and explicit linguistic knowledge. *Studies in Second Language Acquisition*, *37*, 457–486.
- Zimmerman, D. W., & Williams, R. H.** (1982). Gain scores in research can be highly reliable. *Journal of Educational Measurement*, *19*(2), 149–154.

---

**Cite this article:** Li S, Ellis R, and Zhu Y (2019). The associations between cognitive ability and L2 development under five different instructional conditions. *Applied Psycholinguistics* *40*, 693–722. <https://doi.org/10.1017/S0142716418000796>