Studies in Second Language Acquisition, **43** (2021), 523–550
doi:10.1017/S0272263120000601

Research Article

IMPLICITNESS AND EXPLICITNESS IN COGNITIVE ABILITIES AND CORRECTIVE FEEDBACK

A DOUBLE DISSOCIATION?

Yucel Yilmaz *

Indiana University

Gisela Granena

Universitat Oberta de Catalunya

Abstract

This aptitude–treatment interaction study investigated the extent to which explicit and implicit cognitive abilities are differentially related to learning outcomes under two corrective feedback conditions. One hundred and thirteen intermediate English learners of Spanish were randomly assigned to an implicit feedback (recast), explicit feedback (explicit correction), or control group after completing tests from two aptitude batteries (High-Level Language Aptitude Battery [Hi-LAB] and LLAMA). Linguistic improvement on noun-adjective gender agreement and Differential Object Marking was assessed using grammaticality judgment and oral production tasks. Results showed that implicit but not explicit abilities were relevant for the acquisition of gender agreement under implicit feedback as measured by grammaticality judgments. In contrast, explicit but not implicit abilities were relevant for the acquisition of object marking under explicit feedback as measured by oral production. These results lent support to a double dissociation, but they also suggested higher-order interaction effects between the type of cognitive ability, outcome measure, and target structure.

INTRODUCTION

Research in second language acquisition (SLA) that investigates interactions as an attempt to provide insights into the nature of the language learning process is scarce,

This work was supported by a Spencer Foundation Small Research Grant #201500053 to Gisela Granena and Yucel Yilmaz. We would like to thank the CASL for providing us with web-delivered versions of the Hi-LAB tests and for scoring and sending us the data, blind of any specific hypothesis about the results. We would also like to acknowledge the invaluable help provided by Diana Arroyo for her assistance during data collection and analysis.

* Correspondence concerning this article should be addressed to Yucel Yilmaz, Department of Second Language Studies, University of Indiana, Ballantine Hall 715, 1020 E. Kirkwood Ave., Bloomington, Indiana 47405. E-mail: yyilmaz@indiana.edu

© *The Author(s), 2021. Published by Cambridge University Press*

https://doi.org/10.1017/S0272263120000601 Published online by Cambridge University Press

despite its important theoretical and practical implications (DeKeyser, 2012). One of the interactions with the most practical importance is the interaction between instructional treatment and language aptitude, understood as those cognitive aspects of an individual that can determine second language (L2) learning success holding all other relevant factors equal (Doughty, 2014). This type of research has practical implications because it can provide evidence in support of the benefits of matching learners' cognitive strengths to type of instruction to optimize L2 learning. Indirectly, this type of research can also inform SLA theory by providing evidence of the mental processes engaged during learning under particular instructional conditions.

One of the most widely investigated instructional interventions in SLA is corrective feedback (CF), which refers to the reactions that L2 learners receive from their interlocutors, indicating that learners' language production is not target like. Many SLA researchers (e.g., Ellis, 1991; Gass & Mackey, 2006; Long, 1996) agree that evidence indicating what is not possible in the target language (i.e., negative evidence), as conveyed through CF, plays a facilitative role in L2 acquisition. The aptitude-treatment interaction (ATI) study we report investigated two types of feedback, implicit feedback (recast) and explicit feedback (explicit correction), and cognitive aptitudes that could moderate the effectiveness of the two CF types differently. These cognitive aptitudes were measured by tests from two aptitude test batteries (The LLAMA Language Aptitude Tests; Meara, 2005, and Hi-LAB; Linck et al., 2013), and they aimed at throwing light on the cognitive processes engaged in learning under implicit and explicit feedback conditions.

THEORETICAL BACKGROUND

COGNITIVE APTITUDES FOR IMPLICIT AND EXPLICIT L2 LEARNING

Identifying and understanding the factors that influence L2 learning and that can predict rate of acquisition and/or long-term achievement (i.e., ultimate attainment) has been one of the main goals of the field of SLA. Among the many factors that explain variation in L2 learning outcomes, language aptitude has attracted the attention of researchers, especially since the 1950s, for being a powerful predictor of success in language learning. "Language aptitude" is a catchall umbrella term that can refer to a broad range of abilities contributing to L2 learning success. In this article, we use the term to refer to *cognitive aptitudes*, to what Doughty (2019) called "cognitive language aptitude," or those cognitive aspects of an individual that are relatively stable and largely determined by genetics and early childhood language learning experience (DeKeyser & Koeth, 2011) and that can determine how successful an L2 learner can be holding all other relevant factors equal (Doughty, 2013, 2014).

Advances in cognitive psychology, particularly in areas such as implicit learning, as well as in language teaching methodology and in understanding how languages are learned (Long & Doughty, 2009), have given rise to a reconceptualization of the notion of language aptitude in the SLA field. Traditional conceptualizations of aptitude are largely based on Carroll's model of aptitude and the Modern Language Aptitude Test battery (MLAT; Carroll & Sapon, 1959) and see aptitude as a combination of domain-specific abilities that rely on explicit cognitive processes. Carroll (1962) listed four components of aptitude (phonetic coding ability, grammatical sensitivity, memory

https://doi.org/10.1017/S0272263120000601 Published online by Cambridge University Press

abilities, and inductive language learning ability), all of which (except for inductive language learning ability) were measured by the MLAT in varying degrees through five subtests. Carroll (1966) further pointed out that "indeed, several parts of the MLAT are explicitly learning tasks, particularly parts 1, 2, and 5" (p. 20), while parts 3 and 4 (Spelling Clues and Words in Sentences) are measuring "linguistic aptitudes" (Parry & Child, 1990, p. 52). Although Carroll did not make any claims that aptitude is about explicit learning, the MLAT measured aptitude using explicit learning tasks, and this has influenced the understanding of the notion of language aptitude since then.

More recent conceptualizations of aptitude, however, include abilities that involve implicit cognitive processes and domain-general cognitive operations. This is the case of the most recent language aptitude test battery, the High-Level Language Aptitude Battery (Hi-LAB; Linck et al., 2013), a battery that measures 11 cognitive and auditory perceptual abilities grouped into six different constructs, several of which were not present in the MLAT. Two of the constructs in the Hi-LAB, long-term memory retrieval and implicit learning, rely on abilities in the domain of implicit cognitive processes. Long-term memory retrieval is measured by means of a semantic priming task showing subconscious meaning associations (i.e., the ability to activate semantic networks in an unconscious way), whereas implicit learning is measured by a serial reaction time task showing subconscious pattern-learning (i.e., the ability to pick up on patterns without needing to think about them consciously).

Following calls for new aptitude constructs in educational and cognitive psychology (Kaufman et al., 2010; Woltz, 2003), Granena (2013) also proposed a distinction between explicit and implicit language aptitude. Explicit language aptitude includes explicit cognitive abilities that involve attention-driven memory processes and that are relevant to learn a language explicitly and intentionally through reasoning, deliberate hypothesis testing, and memorization. For example, analytic ability, rote memory, and working memory all engage explicit cognitive processes. By contrast, implicit language aptitude includes implicit cognitive abilities that involve cognitive processes labeled in the literature as automatic, associative, sensitive to statistical covariation, nonconscious, and unintentional. It includes abilities such as implicit inductive learning and implicit memory, which are relevant to learn a language through exposure by acquiring patterns unintentionally. Implicit cognitive abilities rely on selective attention, but, unlike explicit abilities, they operate primarily outside of awareness and place minimal demands on central executive resources.

Distinguishing between these two broad types of cognitive aptitudes has several advantages. First, cognitive abilities in the domain of implicit cognition typically show minimal overlap with abilities in the domain of explicit cognition, as suggested by the weak-to-zero correlations between explicit cognitive abilities such as working memory or psychometric intelligence and implicit memory measures of priming or implicit induction (Engle et al., 1999; Kaufman et al., 2010; Kyllonen, 1996; Kyllonen & Christal, 1990; Woltz, 2003). Second, the implicit–explicit dichotomy is a recurrent theme in the field of SLA, going back to, at least in modern times, Krashen's (1981) distinction between acquired (implicit) and learned (explicit) knowledge. Despite the limitations of such a two-system view given the complexity of language learning (Hulstijn, 2015), investigating the relationship between explicit/implicit aptitude and explicit/implicit learning conditions can provide valuable information regarding the learning processes responsible

https://doi.org/10.1017/S0272263120000601 Published online by Cambridge University Press

for the learning outcomes observed (DeKeyser, 2012). This is the goal of ATI (Cronbach & Snow, 1977) research, which examines how learning outcomes depend on the match between specific aptitudes and the treatment received. The effect of the treatment is optimal when a treatment and an individual's aptitude are matched. This allows for the formulation of double dissociation hypotheses stating that learning under explicit instructional conditions will be related to explicit but not implicit cognitive abilities, whereas learning under implicit instructional conditions will be related to implicit but not explicit cognitive abilities. Dissociations are used to infer the existence of separate mental processes, an underlying mental function required by A but not by B and, conversely, a mental function required by B but not by A. Explicit instructional conditions are instructional interventions that include (a) metalinguistic information in the form of rules or metalinguistic terminology or (b) statements directly asking learners to attend to forms or directly informing learners about the inaccuracy of their language production, whereas implicit instructional conditions are instructional interventions including neither (a) nor (b) (Norris & Ortega, 2000).

There is some evidence in support of a single dissociation between implicit and explicit cognitive abilities in learning under explicit instructional conditions (Granena & Yilmaz, 2018). Seven out of the nine studies reviewed in Granena and Yilmaz (2018) showed a significant relationship between learning outcomes under explicit instructional conditions and individual differences in explicit cognitive abilities in at least one of the posttest times and outcome measures. However, the two studies that investigated implicit cognitive abilities reported no significant relationships between implicit cognitive ability and learning outcomes under the explicit condition. The overall correlation across posttests and measurement types was positive and moderate ($r = 0.47$) between learning outcomes under explicit instructional conditions and explicit cognitive abilities, and negative and weak ($r = -0.17$) between outcomes under explicit instructional conditions and implicit cognitive abilities.

However, the evidence in support of the dissociation between these two types of abilities under implicit instructional conditions is not strong (Granena & Yilmaz, 2018). These are conditions that do not include metalinguistic information or instructions to attend to particular forms. Granena and Yilmaz (2019) found that implicit learning ability, as measured by a serial reaction time task, was significantly related to the implicit feedback group's posttest performance after controlling for pretest scores. However, this relationship was limited to one of the two target structures investigated (i.e., gender agreement) and to those items measuring feminine, not masculine, gender agreement. Granena and Yilmaz (2018) argued that very few studies to date had investigated the relative role of both implicit and explicit cognitive abilities in learning under implicit instructional conditions and that, therefore, more research was needed to examine the relationship between implicit cognitive abilities and the effects of implicit instruction.

In general, research on cognitive aptitudes in the field of SLA is not characterized by designs that aim at investigating double dissociations from the point of view of implicit and explicit learning. Research designs tend to include either a single L2 learning condition or a single type of cognitive ability. There are studies, however, that have looked at both types of cognitive abilities, implicit and explicit, conceptualizing them as declarative and procedural memory abilities (Hamrick, 2015; Morgan-Short et al., 2014, 2015; Tagarelli et al., 2016). Hamrick (2015) investigated whether individual differences

https://doi.org/10.1017/S0272263120000601 Published online by Cambridge University Press

in declarative and/or procedural memory abilities predicted the learning of novel syntactic structures in a semiartificial language paradigm under incidental conditions. In this paradigm, words in English (participants' native language) are placed into the syntactic structures of another language, in this case, a language adapted from Persian, to avoid the need for vocabulary pretraining and minimize intentional learning by focusing participants' attention on meaning comprehension of scrambled sentences. The serial reaction time task was used as a measure of procedural memory ability. The results showed that immediate recognition task scores were significantly correlated with declarative memory ability (measured in the study using the LLAMA-B subtest), but not with procedural memory ability. Delayed recognition task scores showed the reverse pattern, that is to say, they were significantly correlated with procedural memory ability, but not with declarative memory ability. This result suggested a relationship between procedural memory ability and the retention of incidentally learned L2 syntax after a period of 1 to 3 weeks of no exposure.

Morgan-Short et al. (2014) used measures of declarative and procedural memory ability to predict individual differences in L2 syntactic development at early and late stages of acquisition under implicit training conditions. Fourteen participants were exposed to an artificial L2 (Brocanto2) without grammatical explanations or instructions to search for rules. The results showed that declarative and procedural memory abilities predicted L2 syntactic development differently. Specifically, there was a positive relationship between declarative memory ability and early L2 syntactic development, and a positive relationship between procedural memory ability and late L2 syntactic development. These results were interpreted as showing that in the early stages of acquisition, learning can occur quickly, relying on the declarative memory system, whereas, in later stages, learning proceeds gradually through repeated exposure relying on the procedural memory system. A follow-up study by Morgan-Short et al. (2015) using neuroimaging techniques and the same implicit training conditions further indicated that neural activity while taking a grammaticality judgment test (GJT) was not related to individual differences in procedural memory ability at the early stage of syntactic development but was related to individual differences in declarative memory ability. However, learners with poorer procedural memory ability obtained poorer GJT scores at the late stage of syntactic development and showed increased levels of neural activation while responding to the GJT. This result was interpreted as showing that learners with poorer procedural memory ability required more effortful, and therefore less efficient, neural recruitment levels.

Tagarelli et al. (2016) predicted that working memory capacity and procedural memory ability, as measured by a serial reaction time task, would be differentially related to semiartificial language learning under incidental and instructed learning conditions. The results were unexpected because working memory was found to be related to learning outcomes under the incidental condition. The study also found a significant negative correlation between incidental learning and procedural memory ability. No relationships were found between learning outcomes and any of the cognitive abilities in the instructed group. Tagarelli et al. argued that the saliency of the pattern could have favored learners' explicit processing under the incidental learning conditions.

Although we see advantages in distinguishing between two broad types of cognitive aptitudes in terms of explicit and implicit cognitive processes, each type of aptitude may not be a unitary construct, but rather, a multifaceted construct with multiple constituents.

https://doi.org/10.1017/S0272263120000601 Published online by Cambridge University Press

Especially in the case of implicit cognitive abilities, the idea of a single, general ability has not been supported in the cognitive psychology literature. Gebauer and Mackintosh (2007), for example, showed that different measures of implicit cognitive ability did not load on a common factor. Specifically, the two major experimental paradigms of implicit learning, artificial grammar and serial reaction time, shared no common variance. Similarly, Kalra et al. (2019) found that a model with four commonly used measures of implicit learning (artificial grammar, probabilistic classification, serial reaction time, and implicit category learning) only revealed good fit when the artificial grammar task was removed from the model. Conway and Christiansen (2005) reported significant differences in implicit sequence learning depending on the modality of stimuli (tactile, visual, or auditory). Finally, Buffington and Morgan-Short (2019) investigated the construct validity of a set of tasks measuring procedural memory. They could not find evidence in support of convergent validity for the tasks. None of the tasks showed statistically significant positive correlations with each other. Tasks included an alternating serial reaction time task, a dual-task version of the weather prediction task, and the Tower of London task. Buffington and Morgan-Short argued that a possible explanation could be that learning abilities in procedural memory are independent of each other because procedural memory encompasses several memory subsystems.

It may be helpful to distinguish between a memory and a learning component in explicit and implicit language aptitude. In practice, there is no strict dividing line between learning and memory, which explains why Buffington and Morgan-Short (2019) claim that the two long-term memory systems (declarative and procedural) support learning and that individual differences in these memory systems should be considered components of aptitude. However, there are reasons to investigate learning and memory separately. Seger (1994), for example, states that "implicit memory involves memory for specific stimuli and implicit learning involves memory for patterns" (p. 165). In the case of implicit cognitive abilities, most research in implicit memory consists of priming studies that measure the effects of stimuli on performance, whereas implicit learning involves performing some inductive process on them to gain new knowledge. Depending on the definition, implicit learning may overlap to a certain extent with procedural memory. This is why Buffington and Morgan-Short (2019) argued that Granena (2013) and Linck et al. (2013) used tasks of procedural memory, in reference to the serial reaction time task. However, in Anderson's (1987) conception of skill learning, procedural memory would not be compatible with Reber's (1993) notion of implicit learning because Anderson's definition requires an initial declarative approach. As a result, some of the tasks proposed to measure procedural learning (or procedural memory acquisition) may not qualify as measures of implicit learning in the Reberian sense.

CORRECTIVE FEEDBACK AND COGNITIVE APTITUDES

CF has been one of the most frequently researched instructional features in relation to cognitive aptitudes. As an example, in Granena and Yilmaz's (2018) research synthesis, 55.6% of the studies that included both an implicit and an explicit instructional condition and at least a cognitive ability were in the area of CF. The interest of the SLA field in CF is also evident in the number of studies that have synthesized research on this topic (Goo

https://doi.org/10.1017/S0272263120000601 Published online by Cambridge University Press

et al., 2015; Li, 2010; Long, 2007; Lyster & Saito, 2010; Mackey & Goo, 2007; Nassaji, 2015; Sheen & Ellis, 2011; Yilmaz, 2016) and that have demonstrated that CF is beneficial for L2 acquisition.

Although there are different ways of classifying feedback types, a widely accepted classification (e.g., Li, 2010) involves determining where along an explicit/implicit continuum feedback types fall. One should note that this classification describes the information provided using the feedback and not the way the learner processes that information. Explicit feedback includes either one or both of the following: (a) metalinguistic information in the form of clues or rules; and (b) information indicating that the learner's production is not targetlike. Feedback types lacking both (a) and (b) have been considered implicit. According to this distinction, explicit correction (i.e., explicit rejection of learners' production, followed by the provision of the targetlike form) and metalinguistic feedback (i.e., comments about the accuracy of learners' nontargetlike production including the provision of metalinguistic terminology, clues, or rules) are classified as explicit. However, recasts (i.e., targetlike reformulations of learners' nontargetlike productions) constitute one type of feedback that is generally considered implicit.[1]

Research has shown that explicit feedback, operationalized as metalinguistic feedback (e.g., Carroll & Swain, 1993; Ellis, 2007; Ellis et al., 2006; Sheen, 2007) or as explicit correction (e.g., Yilmaz, 2012, 2013b), is more effective than implicit feedback, operationalized as recasts. A factor that seems to be behind the greater effectiveness of explicit feedback types is the fact that explicit feedback makes the corrective function more salient, which helps the learner's interpretation and the shift of attention from meaning to form (Carroll, 2001).

Feedback type is one of the main factors that can moderate the effectiveness of feedback as an independent variable. Another important independent variable that can moderate the effectiveness of feedback by interacting with feedback type is cognitive ability. A growing number of studies has investigated the relationship between different cognitive factors and the effectiveness of CF. Most of these studies have focused on language analytic ability (Arroyo & Yilmaz, 2017; Li, 2013; Sheen, 2007; Trofimovich et al., 2007; Yilmaz, 2013a) and working memory capacity or phonological short-term memory (Goo, 2012, 2016; Granena, 2013; Li, 2013; Révész, 2012; Trofimovich et al., 2007). Other variables, such as attention control (Trofimovich et al., 2007), phonetic coding ability (Yilmaz & Koylu, 2016) and explicit language aptitude as measured by the LLAMA subtests B, E, and F, largely based on the MLAT (e.g., Yilmaz & Granena, 2016) have also been investigated, but to a lesser extent. Out of these studies, only Goo (2012, 2016), Li (2013), Sheen (2007), and Yilmaz (2013a) included designs with both implicit and explicit feedback conditions and, therefore, were able to examine any differential relationships between cognitive factors and the effectiveness of different types of feedback.

Li (2013), Sheen (2007), and Yilmaz (2013a) investigated language analytic ability, which can be defined as "the capacity to infer rules of language and make linguistic generalizations or extrapolations" (Skehan, 1998, p. 204). Li (2013) found that analytic ability predicted learning gains in the implicit feedback condition, whereas Sheen (2007) and Yilmaz (2013a) did not. In the explicit feedback condition, Sheen (2007) and Yilmaz (2013a) reported a significant relationship between learning outcomes and analytic ability, while Li (2013) did not. Goo (2012, 2016), Li (2013), and Yilmaz (2013a) investigated working memory, or the capacity to hold information briefly in memory

https://doi.org/10.1017/S0272263120000601 Published online by Cambridge University Press

while performing other mental operations (Baddeley & Hitch, 1974; Kane & Engle, 2002) and the results were also mixed. In the explicit feedback condition, Li (2013) and Yilmaz (2013a) found a statistical relationship between learning outcomes and working memory, whereas Goo (2012, 2016) did not. In the implicit feedback condition, Goo (2012) found a relationship between working memory and learning outcomes, but Goo (2016), Li (2013), and Yilmaz (2013a) did not. Finally, Yilmaz and Granena (2016) investigated explicit language aptitude and found that it was predictive of learning outcomes in the explicit feedback group, but not in the implicit feedback group, a differential effect suggesting that explicit and implicit feedback engage qualitatively different cognitive processes.

As can be seen in this brief review of research, studies on CF and cognitive aptitudes have yielded mixed findings. Given the methodological heterogeneity of the studies regarding issues such as the operationalization of the different types of feedback, target structures, and the measurement of the cognitive variables, it is a difficult, if not impossible, task to account for such mixed results. However, there seems to be a trend showing that in explicit feedback conditions, operationalized as explicit correction (Yilmaz, 2013a; Yilmaz & Granena, 2016) or as metalinguistic information with (Li, 2013; Sheen, 2007) or without (Goo, 2012, 2016) the reformulation of the learner's nontargetlike utterance, learning outcomes tend to be related to an ability in the explicit cognitive domain. This role of explicit cognitive abilities in moderating the effectiveness of explicit feedback may be related to the fact that explicit feedback directly leads learners to search for rules or to formulate and test hypotheses, engaging them in problem-solving and language analysis. Less is known about the role of implicit cognitive abilities due to the lack of studies investigating this type of abilities, especially studies with designs including an explicit and an implicit instructional condition and explicit and implicit cognitive abilities hypothesized to be relevant under each of the treatment conditions.

The current study was part of a larger project investigating the relative effectiveness of two instructional interventions (implicit and explicit CF) as a function of individual differences on cognitive aptitudes for implicit and explicit learning. By including two types of CF and cognitive aptitudes hypothesized to moderate the effectiveness of the two CF types differently, we aimed at investigating possible dissociations between instructional condition and cognitive ability. Previous journal publications from this project reported on partial results by focusing on CF types and a single cognitive ability (Granena & Yilmaz, 2019) or on CF types and multiple individual cognitive abilities and their effects on a single target structure (Yilmaz & Granena, 2019). Finally, a study by Granena (2019) investigated the underlying structure of the battery of cognitive tests reported in this article with all the participants that took part in the first stage of the broader project ($N = 135$). The results showed that the tests loaded onto three different aptitude components, which were interpreted as Explicit Aptitude, Implicit Memory Ability, and Implicit Learning Ability. Explicit aptitude included loadings of tests measuring explicit, attention-driven cognitive abilities that involved conscious and reflective learning processes. Implicit memory ability included loadings of tests measuring unconscious (i.e., effortless and incidental) retrieval of information that was acquired either intentionally or incidentally. Implicit learning ability included the loading of a test measuring implicit inductive learning ability, the ability to learn a pattern or rule through exposure and without the intent to learn the pattern.

The study reported in this article differs from the studies previously published in that it looked at the differential effect of these different aptitude components on the effectiveness

https://doi.org/10.1017/S0272263120000601 Published online by Cambridge University Press

of implicit and explicit CF on two target structures. This approach allowed investigating potential dissociations between type of instructional condition and type of cognitive ability in an attempt to provide insights into the nature of the language learning process when different structures are involved.

The following research question guided this study: Are there any differential effects of different cognitive aptitudes on the effectiveness of feedback types?

METHODOLOGY

BROADER RESEARCH PROJECT AND PARTICIPANTS

The broader project of this study employed a pretest−posttest randomized block design with a control group. It included two stages. Tests of cognitive individual differences and proficiency were administered in the first stage, in which 135 Spanish learners participated. A priori power analysis was not conducted. Rather, we established sample size based on practical considerations such as the study's time frame, number of data-collection sessions involved, and available budget. The first language (L1) of the participants was English, and they had taken two semesters of college-level Spanish. At the end of the first data collection stage, these 135 learners were invited to participate in the second stage of the experiment and were randomly assigned to three groups (implicit feedback, explicit feedback, and control).

In the second stage, a pretest, two treatment sessions, and a posttest were administered to all participants. One hundred and fourteen learners participated in this stage. Because of missing data and the removal of extreme values and high pretest scorers from the data (see Data Analysis for details), various subsets of these participants were used in the current study. In the analysis including the highest number of participants, there were 113 participants (73 female and 40 male) distributed as follows: 42 in the control, 35 in the implicit feedback group, and 36 in the explicit feedback group. The average age of the participants was 20.21 (SD = 4.42). Thirty-three participants (control = 14; implicit = 9; explicit = 10) had studied a different language before (e.g., Chinese, Czech, French, German, Hebrew, Hindi, Italian, Japanese, Portuguese, and Thai), and their self-reported proficiency on these languages ranged from beginner to intermediate.

We administered a cloze test and an oral picture description test[2] as measures of proficiency to compare the groups in their proficiency level at the beginning of the study. One-way ANOVAs carried out to compare cloze and oral picture description test scores (see Appendix B in Online Supplementary Materials for more information about the tasks and descriptive statistics) revealed no differences between the groups (cloze, $F(2, 110) = 1.140$, $p = 0.324$, $\eta^2 = 0.020$; oral picture description, $F(2, 110) = 0.525$, $p = 0.593$, $\eta^2 = 0.009$), indicating that the three groups' proficiency levels were comparable at the beginning of the study.

TARGET STRUCTURES

Two linguistic targets were included in this study: Spanish noun-adjective gender agreement and Differential Object Marking (DOM). The acquisition of these structures has been considered challenging for English speakers. The literature on the acquisition of

https://doi.org/10.1017/S0272263120000601 Published online by Cambridge University Press

Spanish gender agreement has shown that not even advanced learners can reach nativelike accuracy on gender agreement, despite the fact that it is highly frequent in the input (e.g., Granena, 2014). It has been argued (e.g., Fernandez-Garcia, 1999; Granena, 2014; Leeman, 2003) that this learning difficulty can be linked to low perceptual salience and lack of communicative value. The Spanish gender system is binary: nouns can be either masculine (M) or feminine (F). Most Spanish nouns are overtly marked for gender either with the masculine ending—*o* or with the feminine ending—*a.* There are also some nouns that are not marked for gender morphologically (e.g., *pared* "wall-F"). In addition, there are deceptively marked nouns (Alarcón, 2010), which are masculine nouns ending in *a* and feminine nouns ending in *o*, as in *planeta* "planet-M" or *mano* "hand-F." Adjectives change their form by taking either the masculine or the feminine ending to match the gender of the noun they describe. The distance between the noun and the adjective describing the noun can be different depending on the syntactic context. The noun and the adjective can be located within the same phrase (intraphrasal agreement; see Examples [1a] and [1b]) or in different phrases (interphrasal agreement; see Examples [2a] and [2b]).

(1) a. *El pueblo bonito* ("The-M village-M beautiful-M")
b. *La casa bonita* ("The-F house-F beautiful-F")
(2) a. *El libro es rojo.* ("The-M book-M is red-M")
b. *La gorra es roja.* ("The-F hat-F is red-F")

DOM is a phenomenon observed in many languages with overt case marking of direct objects. In these languages, some, but not all, objects are marked because they carry certain semantic and pragmatic features. Spanish DOM is a complex phenomenon where animacy, specificity, properties of the direct object, and the lexical semantics of the verb play a role in determining whether a direct object should be marked. When direct objects are animate, specific, and definite, the preposition "a" obligatorily precedes them (see Examples 3a and 3b). In our treatment and test materials, we only varied animacy and held constant factors including specificity and properties of the direct object to be able to present learners with unambiguous linguistic contexts where the marking of the direct object was either grammatical or ungrammatical. DOM is considered acquisitionally challenging because the preposition "a" has low perceptual salience and performs multiple functions (e.g., locative or directional preposition, dative marker before indirect objects, and dative marker before dative experiencer subjects; Montrul & Gürel, 2015). In addition, it has been shown (e.g., Guijarro-Fuentes, 2012) that the acquisition of DOM is especially difficult for learners whose L1 is a non-DOM language (e.g., English).

(3) a. *Juan besó a la niña.* ("John kissed DOM the girl")
b. *Juan besó la fotografía.* ("John kissed the picture")

TREATMENT TASKS

Each participant carried out two one-way information gap tasks with a native speaker of Spanish (i.e., the experimenter) in two different sessions. These tasks shared the same instructions and content (i.e., pictures), but the content was presented in different

https://doi.org/10.1017/S0272263120000601 Published online by Cambridge University Press

randomized orders. Two sets of PowerPoint slides, one for the participant and one for the experimenter, were used in each task to create contexts for the use of the two target structures. The participants were told to describe the picture on their slides to help the experimenter choose the same picture from a set of three. The slides for gender agreement depicted two versions of an object contrasting with respect to their color, shape, or size. The DOM slides included two scenes that differed regarding the agent and the patient of the action of the verb. In each DOM slide, the verb that needs to be used, the meaning of the verb, and the names of the human characters (if the item designed to elicit DOM in the context of a proper noun) were also provided. Each slide set included 40 slides, 16 for gender agreement and 16 for DOM, and 8 distractors. The slides eliciting gender agreement included only overt gender nouns (e.g., *el castillo*) and adjectives that carry gender marking (e.g., *amarillo/a*). The nouns were balanced for type of gender. The slides targeting DOM had the following characteristics: (a) direct objects were balanced for animacy (animate and inanimate); (b) animate nouns were balanced in terms of proper/common status (*Maria* vs. *la/una profesora*); and (c) all direct objects were feminine nouns to avoid having to use the Spanish contraction between the direct object marker *a* and the masculine article *el* (i.e., *al*). Pilot testing indicated that the slides were effective in creating contexts for the use of the target structures.

Feedback Treatment

During the treatment tasks, learners' errors on the two target structures were treated depending on learners' group assignments. The control group did not receive any feedback but completed the treatment tasks with the experimenter. The implicit feedback group received partial recasts. These recasts only reformulate the erroneous segment of the learner's utterance. In the case of feedback focusing on gender agreement, the reformulated segment typically included the adjective and the noun. To avoid reformulating errors that were not the target of the study (e.g., determiners), the reformulation did not include determiners. However, other linguistic elements that intervene between the noun and the adjective, such as the copula *está* (see Episode 1), were repeated as they were in the learner's utterance regardless of whether they were targetlike. We thought that dropping them from the reformulation could decrease the detectability of the negative evidence by decreasing the similarity between the original utterance and its reformulated version. The recasts focusing on DOM also reformulated only the erroneous segment, which included the verb and the direct object with its determiner in cases in which the use of the preposition—*a* is ungrammatical, and it included the verb, the proposition, and the direct object with its determiner in cases in which the use of the preposition is grammatical. The verb was included in the reformulation because we thought that in cases in which the learners oversupplied the preposition, a reformulation consisting of only the direct object with its determiner could be ambiguous.

Episode 1: Implicit feedback on a gender agreement error
Learner: **La* ***pelota*** *está* ***viejo***.
"The ball is old."
Experimenter: ***pelota*** *está* ***vieja***, *siguiente*?
"ball is old, next?"

https://doi.org/10.1017/S0272263120000601 Published online by Cambridge University Press

Episode 2: Implicit feedback on a DOM error
Learner: **Pablo toca* ***Gloria***.
"Pablo is touching Gloria."
Experimenter: *toca a Gloria, siguiente*?
"is touching Gloria, next?"

The explicit group received explicit corrections, which differed from the reformulations received by the implicit group with respect to the language that preceded the feedback: *No es correcto; tienes que decir …*" ("It is not correct; you should say …"). This fixed language formula served as direct rejections of learners' utterances and explicit reformulations of their errors. The reformulations in the explicit correction were also partial and provided in the same way as the reformulations in the implicit group.

In both feedback conditions, the experimenter prevented learners from repairing their original nontargetlike utterances in their turn after the feedback by using the information provided in the feedback, a phenomenon known as repair or modified output. The experimenter blocked the repair opportunities by directing the learners' attention to the remaining parts of the task after the feedback, using words or phrases such as "*qué más?/ siguiente*?" ("what else?/next?") with interrogative intonation. The reason why we chose to block opportunities for modified output was that allowing modified output might introduce an additional source of variability that can preclude one from attributing the results of the study to the feedback factor alone. Independent-samples *t*-tests were performed on each target separately to confirm whether the feedback groups received an equivalent amount of feedback. This analysis showed that the feedback groups were equivalent in the amount of feedback on gender agreement ($t(69) = 1.227$, $p = 0.224$, $d = 0.291$), however, the implicit feedback group received significantly more feedback on DOM than the explicit feedback group ($t(70) = 3.068$, $p = 0.003$, $d = 0.723$) (see Appendix B in Online Supplementary Materials for descriptive statistics).

Episode 3: Explicit feedback on a gender agreement error
Learner: **El* ***pelota sucio*** *es en la caja.*
"The dirty ball is in the box."
Experimenter: *No es correcto; tienes que decir "**pelota sucia**," siguiente?*
"It is not correct. You should say 'dirty ball,' next?"

Episode 4: Explicit feedback on a DOM error
Learner: **La niña besa* ***un mujer***.
"The girl kisses a woman."
Experimenter: *No es correcto; tienes que decir besa "**a un mujer**," siguiente?*
"It is not correct. You should say 'kisses a woman,' next?"

Cognitive Measures

Eight computer-based cognitive tests, four from the LLAMA aptitude test battery (Meara, 2005) and four from the Hi-LAB (Linck et al., 2013), were included in this study. We briefly describe the tests in the following text (for more information on the tests and test scoring, see Granena, 2019).

https://doi.org/10.1017/S0272263120000601 Published online by Cambridge University Press

LLAMA-B. This test measures the ability to learn new vocabulary. The 20 words to be learned, taken from a Central American language, are presented visually and linked to a target image. Participants are given 2 minutes to study the word-picture associations. In the testing phase, the program displays one of the words, and participants have to identify the correct picture on the screen. The program calculates an accuracy score out of a maximum of 100. The internal consistency of the test, as indexed by Cronbach's alpha, was 0.81.

LLAMA-D. This test measures the ability to recognize sound sequences. Participants are asked to listen to 10 words in a language with which they are not familiar. Next, participants are tested on the extent to which they can recognize the words they heard in the presentation phase by discriminating between old and novel items. The program computes a score out of a maximum of 75. The internal consistency of the test, as indexed by Cronbach's alpha, was 0.50.[3]

LLAMA-E. This test measures the ability to form sound-symbol associations. It requires participants to work out relationships between sounds (i.e., recorded syllables), 24 in total, and written representations of those sounds in an unfamiliar alphabet. Participants are given 2 minutes to study the associations. In the testing phase, participants hear items, including two syllables, and make a choice between two possible symbol representations. A score is automatically computed out of a maximum of 100. The internal consistency of the test, as indexed by Cronbach's alpha, was 0.72.

LLAMA-F. This test measures the ability to infer the rules of an unknown language. Participants see a set of pictures and sentences in an unfamiliar language describing the pictures (20 in total) and are asked to work out the rules that operate in the language. In the testing phase, participants see a picture and two sentences and have to choose the sentence they consider correct. The program automatically calculates an accuracy score out of a maximum of 100. The internal consistency of the test, as indexed by Cronbach's alpha, was 0.69.

Paired associates (Hi-LAB). This test is a measure of explicit associative memory. Participants are asked to study 20 word pairs, each composed of an English word and a nonword, and tested on the extent to which they remember these word pairs. An accuracy score was derived by summing the total number of correctly recalled English words. The internal consistency of the test, as indexed by Cronbach's alpha, was .87.

Letter span test (Hi-LAB). This test is a measure of phonological short-term memory. Participants see 21 lists of letters of lengths from three to nine on the screen and are asked to recall them in the order they were presented. An accuracy score was derived by summing the number of letters participants were able to recall in the right order. The internal consistency of the test, as indexed by Cronbach's alpha, was 0.76.

Available long-term memory synonym test (Hi-LAB). Available Long-Term Memory (ALTM) is a test of associative priming and measures the extent to which the activation of one concept in participants' lexicon leads to the activation of other related concepts. The

https://doi.org/10.1017/S0272263120000601 Published online by Cambridge University Press

test is composed of two subtasks. The priming task involves listening to a list of five words and deciding which of the two words that are shown afterward have more synonyms on the list. The comparison task involves deciding whether the words shown in pairs have a different or similar meaning. Eighteen sets of 12-word pairs are presented in the comparison task, each of which comes after a priming list from the priming task. Nine of the sets are primed, meaning that one or both words in the pairs are synonyms of one of the two topic words from the priming list. Nine sets are unprimed, which means that none of the words is synonyms of the two topic words from the preceding priming list. A residual priming score was calculated by regressing rate scores (number of correct responses divided by the amount of time taken for responses) for the primed sets onto rate scores for the unprimed sets (for further details about this task and the scoring method, see Linck et al., 2013). The internal consistency of the test, as indexed by Cronbach's alpha, was 0.70.

Serial reaction time test (Hi-LAB). This test is a measure of implicit sequence learning. Participants are presented with four boxes displayed horizontally on the screen, and an asterisk appears in one of these boxes. Participants are asked to press the button that corresponds to the location of the asterisk on their keyboard. There are six blocks of 96 trials. In the first and last blocks, the asterisk does not appear in a particular order, whereas, in the rest of the blocks, the asterisk appears in a repeating sequence of length 12. A facilitation score was calculated by subtracting the median reaction times in the last sequential block from the median reaction times in the last random block (Linck et al., 2013). The reliability of the test was 0.79, using the split-half method.

LANGUAGE OUTCOME MEASURES

Learners' linguistic knowledge of gender agreement and DOM were measured through two versions of an oral production task (OPT) and two versions of a GJT. The two tests included completely different items. The two test versions of each test type also included completely different items and were counterbalanced within each feedback condition. Half the items targeting the same linguistic structure were repeated from the treatment tasks,[4] and half appeared only on the outcome measures (i.e., novel). The status of the items as to whether they were novel versus repeated was determined randomly. DOM items were balanced for animacy (animate or inanimate), and DOM animate items were further balanced as to whether they included a proper or common noun. Items targeting gender agreement were balanced for phrasal context (interphrasal or intraphrasal), but not for gender. Gender items were arranged such that there were more feminine nouns (6 in the OPT; 10 in the GJT) than masculine nouns (four in the OPT; six in the GJT). This decision was based on previous research indicating that the feminine is the marked[5] form (Finneman, 1992).

The OPT involved describing the difference between sets of two pictures. The OPT was included in the study to measure learners' knowledge of the target form that could be deployed under conditions that require attention to be primarily on meaning. The task included 16 contexts for the use of each linguistic target and 8 contexts to distract learners' attention away from the target forms. Pilot testing revealed that the test was successful in eliciting the target forms. A gender agreement score was computed by dividing the number of accurate responses in obligatory contexts by the total number of obligatory contexts. An obligatory context for gender agreement was defined as an instance in which

https://doi.org/10.1017/S0272263120000601 Published online by Cambridge University Press

a gender-showing adjective occurred in the same utterance as the noun it described. A DOM score was computed by dividing the number of accurate responses in obligatory contexts by the total number of relevant (obligatory plus nonobligatory) contexts. An utterance that included a subject, a verb, and an animate direct object was considered an obligatory context for DOM, whereas an utterance including a subject, a verb, and an inanimate direct object was considered a nonobligatory context. Interrater reliability was assessed by having two native-speakers of Spanish code 10% of the data independently. The percentage agreement between the two coders was 96% for learners' accuracy in relevant contexts. After the coding, disagreements were discussed and resolved.

The instructions of the GJT required learners to determine whether the sentences presented in the items were grammatical and to provide a written correction for the ones they considered ungrammatical. The GJT was included in the study to measure learners' knowledge of the target form under conditions that demand focal attention to be primarily on the linguistic form. The test was untimed. There were 48 items in the GJT (16 on each linguistic target and 16 distractors) balanced for grammaticality (eight grammatical and eight ungrammatical items for each structure). Learners' responses to GJT items were coded as correct or incorrect, and an accuracy score was calculated for grammatical and ungrammatical items separately based on the ratio of correct responses to total items. For a GJT item to be coded as correct, the response should have accurately identified sentences as grammatical or ungrammatical and included the relevant correction if the item was ungrammatical. Reliability coefficients were computed for the grammatical and ungrammatical items separately for each linguistic target and for each test version using Cronbach's alpha. It was found that the reliability coefficients for grammatical items were much lower (DOM, version A = 0.35, version B = 0.37; Gender, version A = 0.30, version B = 0.34) than the reliability coefficients for ungrammatical items (DOM, version A = 0.76, version B = 0.79; Gender, version A = 0.82, version B = 0.83), suggesting that grammatical items would not discriminate well among learners. Based on this conclusion, we decided not to use grammatical GJT scores in our statistical analyses.

The data from learners who scored above 0.80 (80%) on the pretest were excluded from the analysis to guard against ceiling effects. Four learners scored above 0.80 in the GJT on ungrammatical gender items. No cases reached the cutoff level in the GJT on ungrammatical DOM items. However, 48 cases in gender OPT and three cases in OPT DOM scored above 0.80 (see Appendix C in Online Supporting Materials for further details).

PROCEDURE

Participants met with the experimenter three times at a research lab in two different stages. Data were collected over 14 months (Stage 1 = 6 months, Stage 2 = 8 months). In the first stage, participants met with the experimenter once and took the cognitive tests and the proficiency tasks. Web-delivered versions of the Hi-LAB (Linck et al., 2013) tests were administered online. Center for Advanced Study of Language (CASL) researchers scored the tests and sent us the scores, blind of any specific hypotheses about the results. In the second stage, learners participated in two data-collection sessions with the experimenter. They took the pretest and performed the first treatment task in the first session. One day later, they carried out the second treatment task and took a posttest immediately after the treatment task. At each testing session (i.e., pretest and posttest), the GJT was administered after the

https://doi.org/10.1017/S0272263120000601 Published online by Cambridge University Press

OPT. At the treatment tasks, the learner and experimenter sat facing each other without visual access to each other's computer screens. A glossary containing a list of the Spanish words used in the experiment and their English translations was available for learners during the treatment sessions and could be consulted by learners on demand.

RESULTS

OVERVIEW OF STATISTICAL PROCEDURES

To answer our research question regarding the interaction between cognitive aptitudes and feedback types, we first ran principal components analysis (PCA) to find an optimal way of combining the eight cognitive measures. The analysis was performed following the procedure described in Granena (2019) for the broader sample in the research project and factor scores were saved as variables. The Kaiser−Meyer−Olkin measure of sampling adequacy (a measure of how suited data are for factor analysis) was 0.72. Conventionally, values that are higher than 0.60 suggest that the proportion of common variance among the variables is low (partial correlations should not be very large if distinct factors are expected to emerge). Bartlett's test of sphericity was significant ($p < 0.001$), indicating that the relationships among the variables were not due to chance. The results of the PCA replicated the results in Granena (2019), revealing three factors with eigenvalues greater than 1.0, which accounted for 58.78% of the total variance. The first component with loadings greater than 0.4 from five tests (Letter Span, Paired Associates, LLAMA-B, LLAMA-E, and LLAMA-F) accounted for 27.48% of the total variance. The second component with loadings greater than 0.4 from two tests (ALTM Synonym and LLAMA-D) accounted for 16.08% of additional variance. Finally, the third component with loadings greater than 0.4 from two tests (Letter Span, a negative loading,[6] and Serial Reaction Time, a much stronger positive loading) accounted for 15.22% of additional variance. Following Granena (2019), the three components were labeled "Explicit Learning Ability" (the ability to learn intentionally, through reasoning, deliberate hypothesis testing, and memorization), "Implicit Memory Ability" (the ability to access information that was intentionally or incidentally acquired information without any conscious effort), and "Implicit Learning Ability" (the ability to learn patterns or rules through exposure and unintentionally).

To answer our research question, we first calculated gain scores for each outcome measure and target structure by subtracting pretest scores from posttest scores. Then, we performed a custom Analysis of Covariance (ANCOVA) model on each gain score type, with Group as the between-subjects factor and the three cognitive aptitudes as covariates. The models also included the interaction terms between the covariates and Group. If a covariate by group interaction was found to be significant, the interaction was probed using two follow-up analyses. First, bivariate correlations between gain scores and each cognitive aptitude for each group were computed.[7] Second, an analysis of simple slopes was carried out, which is automatically provided by the Process macro (version 3.4) developed for SPSS by Hayes (2017) as part of its regression-based moderation analysis. The simple slopes analysis involves determining a point on the distribution of the moderator variable, which is the cognitive aptitude in our case, and estimating whether the groups' averages differ, conditioned on the value of the cognitive aptitude (Hayes & Montoya, 2017). In these regression models, gain scores were the outcome variable. Two

https://doi.org/10.1017/S0272263120000601 Published online by Cambridge University Press

dummy-coded variables representing the levels of the feedback variable, the cognitive aptitude, and two interaction terms between the cognitive aptitude and each dummy-coded variable were the predictor variables.

To determine the extent to which data conformed to normality, we inspected the kurtosis and skewness values of all scores, using the range between 2 and –2 as the target range (George & Mallery, 2010). All values fell within the range except for pretest GJT scores for DOM, where the kurtosis value was unusually high (Pretest GJT DOM, skewness = 2.78, kurtosis = 10.23). Pretest GJT DOM scores were screened for outliers using a boxplot, and one extreme value that fell outside the 1.5 interquartile range was detected and removed from the analysis. As a result, the kurtosis value decreased substantially (3.96), and we decided that this new value did not constitute a severe departure from normality. The assumption of homogeneity of variance was checked by calculating a variance ratio, which is defined as the ratio of the largest variance to the smallest variance of the groups. The ratios were within the recommended boundary of 3.0 (Dean & Voss, 1999), indicating that the assumption of homogeneity of variance was met. In the following text, we report the results of the statistical analyses for each target structure and outcome measure separately.

GENDER AGREEMENT

The first analysis focused on GJT gender agreement scores (see Table 1 for descriptive statistics). Before running the ANCOVA model including all three cognitive aptitudes, we confirmed that the three groups were comparable on each of the aptitude scores (see Table 2 for descriptive statistics) using one-way ANOVAs (see Appendix D in Online Supporting Materials for details). The custom ANCOVA model did not reveal any

TABLE 1. Descriptive statistics for gender agreement test scores.

			Pretest			Posttest			Gain		
Group	Test	*N*	*M*	*SD*	95% CI	*M*	*SD*	95% CI	*M*	*SD*	95% CI
Control	GJT	38	0.17	0.23	0.10, 0.25	0.28	0.29	0.18, 0.37	0.11	0.28	0.01, 0.20
Implicit		35	0.27	0.23	0.19, 0.35	0.44	0.33	0.32, 0.55	0.17	0.28	0.07, 0.27
Explicit		35	0.23	0.27	0.13, 0.32	0.36	0.34	0.24, 0.48	0.14	0.25	0.05, 0.22
Control	OPT	21	0.63	0.12	0.58, 0.68	0.70	0.20	0.61, 0.79	0.07	0.19	−0.01, 0.16
Implicit		24	0.65	0.11	0.60, 0.69	0.83	0.13	0.77, 0.88	0.18	0.12	0.13, 0.23
Explicit		21	0.60	0.13	0.54, 0.66	0.83	0.15	0.76, 0.90	0.23	0.17	0.15, 0.31

TABLE 2. Descriptive statistics for cognitive aptitude scores.

	Explicit learning ability			Implicit memory ability			Implicit learning ability		
	M	*SD*	95.0% CI	*M*	*SD*	95.0% CI	*M*	*SD*	95.0% CI
Control	0.17	1.02	−0.18; 0.51	0.06	1.10	0.31; 0.43	−0.02	0.98	−0.35; 0.30
Implicit	−0.05	0.98	−0.38; 0.28	−0.02	0.99	−0.36; 0.31	−0.13	1.00	−0.47; 0.21
Explicit	−0.13	1.00	−0.49; 0.22	−0.04	0.92	−0.37; 0.29	0.17	1.03	−0.20; 0.53

https://doi.org/10.1017/S0272263120000601 Published online by Cambridge University Press

significant main effect for group ($F(2, 91) = 1.282$, $p = 0.282$, $\eta^2 = 0.023$) or for any of the cognitive aptitudes (Explicit Learning Ability, $F(1, 91) = 3.286$, $p = 0.073$, $\eta^2 = 0.029$; Implicit Memory Ability, $F(1, 91) = 0.975$, $p = 0.326$, $\eta^2 = 0.008$; Implicit Learning Ability, $F(1, 91) = 1.597$, $p = 0.210$, $\eta^2 = 0.01$). The interaction between Implicit Memory Ability and Group ($F(2, 91) = 0.919$, $p = 0.403$, $\eta^2 = 0.016$) as well as the interaction between Explicit Learning Ability and group ($F(2, 91) = .135$, $p = 0.874$, $\eta^2 = 0.002$) were not significant. However, the interaction between Implicit Learning Ability and Group was significant, $F(2, 91) = 4.288$, $p = 0.017$, $\eta^2 = 0.077$.

The significant interaction effect indicates that the relationship between Implicit Learning Ability and GJT gender gain scores changed depending on Group. Follow-up bivariate correlations (see Table 3) revealed that Implicit Learning Ability correlated significantly only with the gains of the implicit feedback group. The strength of this relationship was moderate.

We also probed this interaction by the analysis of simple slopes, which involved running two regression models using two different reference categories in our dummy coding to carry out tests for all possible contrasts between the groups (implicit vs. control, explicit vs. control, and explicit vs. implicit). The results of this analysis are shown in Table 4 (see Appendix E in Online Supplementary Materials for more information about the regression models). At 1 SD below and at the mean of Implicit Learning Ability, there were no significant differences between the groups. At 1 SD above the mean of Implicit Learning Ability, the implicit feedback group performed significantly better than both the explicit feedback and control groups, but there was no significant difference between the explicit feedback and control group.

TABLE 3. Correlations between implicit learning ability and GJT gender agreement scores.

	r	*p*
Control	0.104	0.554
Implicit feedback	0.435	0.009
Explicit feedback	−0.262	0.142

TABLE 4. Results of simple slopes analysis: Implicit learning ability X GJT gender scores.

		df	*b*	*t*	*p*	*d*
1 SD Below	Implicit vs. control	97	−0.005	−0.060	0.952	0.012
	Explicit vs. control	97	0.108	1.204	0.231	0.244
	Implicit vs. explicit	97	−0.113	1.299	0.197	0.263
Mean	Implicit vs. control	97	0.087	1.467	0.147	0.297
	Explicit vs. control	97	0.026	0.42	0.675	0.085
	Implicit vs. explicit	97	0.062	1.013	0.313	0.206
1 SD Above	Implicit vs. control	97	0.18	2.034	0.044	0.413
	Explicit vs. control	97	−0.058	−0.681	0.497	0.138
	Implicit vs. explicit	97	0.234	2.768	0.007	0.562

https://doi.org/10.1017/S0272263120000601 Published online by Cambridge University Press

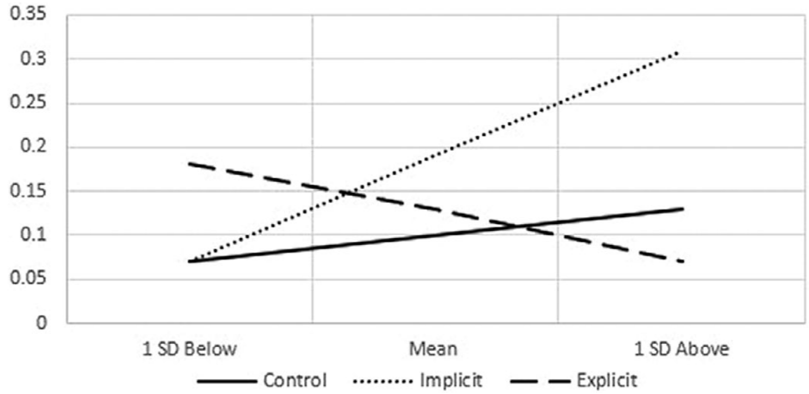

FIGURE 1. Interaction between Feedback Group and Implicit Learning Ability

The interaction plot in Figure 1 displays the nature of the interaction between Group and Implicit Learning Ability. As shown by the slopes for different groups, the differences between the groups are minimal at 1 SD below the mean and the mean, but the gap between the scores of the implicit feedback group and the other two groups widens at 1 SD above the mean.

The second group of analysis was carried out on OPT gender agreement scores (see Table 1 for descriptive statistics). The results of the custom ANCOVA model showed a significant Group effect ($F(2, 50) = 6.868$, $p < 0.001$, $\eta^2 = 0.226$). There were no significant main effects for the covariates (Explicit Learning Ability, $F(1, 50) = 2.001$, $p = 0.163$, $\eta^2 = 0.029$; Implicit Memory Ability, $F(1, 50) = 0.024$, $p = 0.879$, $\eta^2 < 0.001$; Implicit Learning Ability, $F(1, 50) = 0.002$, $p = 0.966$, $\eta^2 < 0.001$) and no significant two-way interactions between the covariates and Group (Explicit Learning Ability × Group, $F(2, 50) = 0.193$, $p = 0.825$, $\eta^2 = 0.005$; Implicit Memory Ability × Group, $F(2, 50) = 0.704$, $p = 0.500$, $\eta^2 = 0.020$; Implicit Learning Ability × Group, $F(2, 50) = 0.310$, $p = 0.735$, $\eta^2 = 0.009$). These results indicate that the cognitive aptitudes measured in this study did not influence the OPT gender scores of the feedback groups differentially.

DOM

The first analysis on DOM targeted the GJT score (see Table 5 for descriptive statistics). The results of the custom ANCOVA model showed a significant Group effect ($F(2, 93) = 7.099$, $p = 0.001$, $\eta^2 = 0.127$). Neither the main effects for the covariates (Explicit Learning Ability, $F(1, 93) = 1.417$, $p = 0.237$, $\eta^2 = 0.012$; Implicit Memory Ability, $F(1, 93) = 1.868$, $p = 0.175$, $\eta^2 = 0.016$; Implicit Learning Ability, $F(1, 93) = 0.001$, $p = 0.975$, $\eta^2 < 0.001$) nor the two-way interactions between the covariates and Group were significant (Explicit Learning Ability × Group, $F(2, 93) = 0.218$, $p = 0.804$, $\eta^2 = 0.003$; Implicit Memory Ability × Group, $F(2, 93) = 0.360$, $p = 0.698$, $\eta^2 < 0.001$; Implicit Learning Ability × Group, $F(2, 93) = 0.034$, $p = 0.967$, $\eta^2 = 0.009$). These results indicate that the

https://doi.org/10.1017/S0272263120000601 Published online by Cambridge University Press

TABLE 5. Descriptive statistics for DOM.

			Pretest			Posttest			Gain		
Group	Test	*N*	*M*	*SD*	95% CI	*M*	*SD*	95% CI	*M*	*SD*	95% CI
Control	GJT	42	0.07	0.14	0.03, 0.12	0.11	0.18	0.05, 0.17	0.04	0.13	0.00, 0.08
Implicit		35	0.06	0.10	0.03, 0.10	0.19	0.25	0.10, 0.27	0.12	0.24	0.04, 0.20
Explicit		36	0.05	0.09	0.02, 0.08	0.27	0.27	0.18, 0.36	0.22	0.23	0.14, 0.30
Control	OPT	40	0.52	0.09	0.49, 0.55	0.57	0.12	0.53, 0.60	0.04	0.12	0.00, 0.08
Implicit		36	0.53	0.12	0.49, 0.56	0.74	0.17	0.68, 0.80	0.22	0.19	0.15, 0.28
Explicit		34	0.51	0.09	0.48, 0.55	0.82	0.16	0.77, 0.88	0.31	0.18	0.25, 0.37

TABLE 6. Correlations between explicit learning ability and OPT DOM scores.

	R	*p*
Control	−0.020	0.980
Implicit feedback	−0.282	0.096
Explicit feedback	0.346	0.057

cognitive aptitudes measured in this study did not influence the GJT DOM scores of the feedback groups differentially.

The second analysis focused on the OPT scores. The results of the custom ANCOVA revealed a significant main effect for Group ($F(2, 91) = 29.876$, $p < 0.001$, $\eta^2 = 0.825$), but no significant main effects for the covariates (Explicit Learning Ability, $F(1, 91) = 0.010$, $p = 0.922$, $\eta^2 < 0.001$; Implicit Memory Ability, $F(1, 91) = 0.006$, $p = 0.938$, $\eta^2 < 0.001$; Implicit Learning Ability, $F(1, 91) = 0.160$, $p = 0.690$, $\eta^2 = 0.004$). The interactions between Implicit Memory Ability and Group ($F(2, 91) = 0.147$, $p = 0.864$, $\eta^2 = 0.004$) and between Implicit Learning Ability and Group ($F(2, 91) = 0.057$, $p = 0.944$, $\eta^2 = 0.004$) were not significant. However, the interaction between Explicit Learning Ability and Group was significant, $F(2, 91) = 3.650$, $p = 0.030$, $\eta^2 = 0.100$, indicating that the relationship between Explicit Learning Ability and OPT gain scores differed across feedback groups.

Given the significant difference between the feedback groups in the amount of feedback on DOM, we performed additional analyses to determine whether feedback amount would influence the interactions between the cognitive aptitudes and Group. We found that there were no discrepancies between the results of these analyses and the results of the analyses reported in the preceding text that did not take feedback amount into consideration (see Appendix B Online Supplementary Materials for the details of these analyses). Based on this result, we proceeded with our analysis probing the interaction between Explicit Learning Ability and Group.

Follow-up bivariate correlations between OPT DOM gain scores and Explicit Learning Ability carried out for each group separately showed no significant correlations (see Table 6), but the correlation between Explicit Learning Ability and the explicit group's OPT gain scores approached the level of significance. This relationship was moderately weak and positive, meaning that the higher the Explicit Learning Ability, the higher the OPT scores.

https://doi.org/10.1017/S0272263120000601 Published online by Cambridge University Press

TABLE 7. Results of simple slopes analysis: explicit learning ability X OPT DOM scores.

		df	*b*	*t*	*p*	*d*
1 SD Below	Implicit vs. control	97	0.223	4.29	<0.001	0.871
	Explicit vs. control	97	0.24	4.547	<0.001	0.923
	Implicit vs. explicit	97	−.016	−0.320	0.75	0.064
Mean	Implicit vs. control	97	0.172	4.807	<0.001	0.976
	Explicit vs. control	97	0.293	7.86	<0.001	1.596
	Implicit vs. explicit	97	−0.121	−3.257	0.015	0.661
1 SD above	Implicit vs. control	97	0.121	2.411	0.018	0.49
	Explicit vs. control	97	0.346	6.629	<0.001	1.346
	Implicit vs. explicit	97	−0.226	−4.143	<0.001	0.841

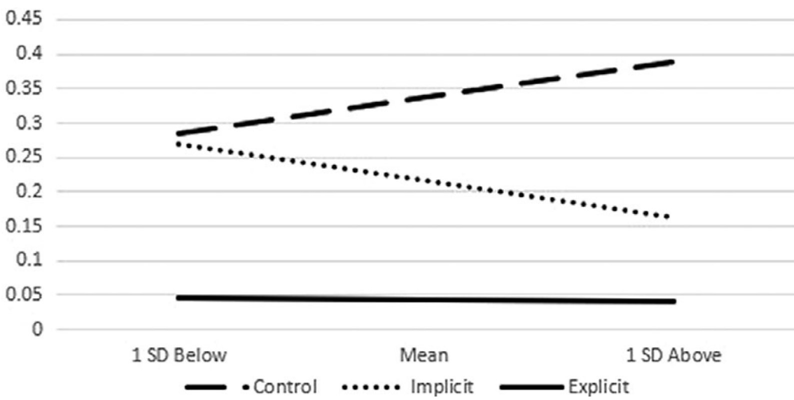

FIGURE 2. Interaction between Feedback Group and Explicit Learning Ability

We also probed this interaction through the analysis of simple slopes. Again, we carried out two regression models, using two different groups as the reference category in our dummy coding to be able to carry out tests for all the possible contrasts between the feedback groups. The results of the analysis are reported in Table 7 (see Appendix E in Online Supplementary Materials for more information). At 1 SD below the mean of Explicit Learning Ability, both the implicit and the explicit group significantly outperformed the control group, but there was no significant difference between the implicit feedback and the explicit feedback group. At the mean and at 1 SD above the mean of Explicit Learning Ability, both the implicit and explicit feedback groups significantly outperformed the control group, and the explicit feedback group significantly outperformed the implicit feedback group.

The interaction plot in Figure 2 confirms these results, providing visual information about the nature of the interaction between Group and Explicit Learning Ability. As can be seen from Figure 2, the slopes for the explicit feedback and implicit feedback groups indicate that the gap between them on OPT scores widens as Explicit Learning Ability

https://doi.org/10.1017/S0272263120000601 Published online by Cambridge University Press

increases. Because the implicit feedback group's gains are in an inverse relationship with Explicit Learning Ability, their gains decrease as their Explicit Learning Ability increases.

DISCUSSION

This study investigated whether a set of different cognitive abilities moderated L2 learning differently under explicit and implicit CF conditions. Our research question asked whether there would be any differential effects of cognitive aptitudes on the effectiveness of feedback types. The eight cognitive measures used in the study loaded onto three components that were labeled as "Explicit Learning Ability," "Implicit Memory Ability," and "Implicit Learning Ability." None of these aptitude components had a moderating effect on OPT gender agreement or GJT DOM gain scores. However, implicit learning ability had a moderating effect on GJT gender agreement gain scores, while explicit learning ability had a moderating effect on OPT DOM gain scores.

Follow-up analyses on these interactions showed that the higher the ability for implicit learning, the higher the GJT gender agreement scores of the implicit feedback group. In addition, a comparison of the groups' GJT gender agreement scores at three levels of implicit learning ability further showed that the groups were not statistically different at 1 SD below the mean or at the mean of implicit learning ability, but the implicit feedback group outperformed the other two groups at 1 SD above the mean of implicit learning ability.

In the case of explicit learning ability, follow-up correlational analyses could not determine the source of the interaction effect because none of the groups' OPT DOM gain scores correlated significantly with this ability. Probing interaction effects using simple slopes analysis was more revealing. This analysis showed that the explicit and implicit feedback groups' OPT DOM gain scores were comparable at 1 SD below the mean of explicit learning ability, but the explicit feedback group significantly outperformed the implicit feedback group at the mean level and at 1 SD above the mean of explicit learning ability.

These results provide partial support for a double dissociation between cognitive aptitudes for implicit and explicit learning and implicit and explicit instructional conditions. That is, under explicit feedback, the ability for explicit learning, not the ability for implicit learning, played a role, whereas, under implicit feedback, the ability for implicit learning, not the ability for explicit learning, played a role. In the following text, we provide possible explanations for the presence (or absence) of relationships in this double dissociation. Learners with high explicit learning ability, who were good at learning intentionally, through attention-driven memory processes, reasoning, and deliberate hypothesis testing, were the ones that benefited the most from explicit feedback. It could be that the directness of the message provided in the explicit feedback regarding the error and how the error should be repaired might have unambiguously conveyed the negative evidence to the learners and put pressure on them to switch attention from meaning to form. As a result, the learners might have intentionally searched for rules, tested hypotheses, and engaged in language analysis. In contrast, cognitive processes that are automatic, nonconscious, and unintentional played no role in the extent to which learners benefited from explicit feedback, perhaps

https://doi.org/10.1017/S0272263120000601 Published online by Cambridge University Press

because these processes were not relevant to interpret explicit feedback and, therefore, did not have an influence on learning outcomes. In the case of learning under implicit feedback, implicit learning ability, or the ability to learn without awareness or unintentionally, had a facilitative role. It is possible that the absence of directness and metalinguistic information in the implicit feedback might have helped those learners with higher implicit learning ability process feedback holistically and in context, without the mediation of analytical and metalinguistic thinking. In addition, the absence of directness and metalinguistic information in the feedback might have obviated the need to engage explicit learning abilities.

Support for the double dissociation was partial because differential relationships were only observed for one of the two outcome measures in each target structure. Contrary to what one would expect, the effect of implicit learning ability was observed in the outcome measure supposed to tap more explicit knowledge (grammaticality judgments), whereas the effect of explicit learning ability was observed in the outcome measure supposed to tap more implicit knowledge (oral production). As pointed out in the previous literature (e.g., Ellis, 2005), there are no pure measures of either knowledge type. Tests cannot be expected to measure a single knowledge type. Therefore, the OPT and GJT used in this study might have tapped the type of knowledge that they are not expected to tap. Further research triangulating learning outcomes with awareness measures providing evidence of the extent to which learners are conscious of the knowledge they gain and the type of L2 processing they engage in during the treatment could (perhaps) throw more light on this complicated puzzle.

Another reason why the support for the double dissociation was partial is that each target structure displayed a different relationship pattern between the aptitude components identified and pretest-posttest improvement. Explicit learning ability, not implicit learning or implicit memory abilities, facilitated the effect of explicit feedback on DOM, whereas implicit learning ability, not implicit memory or explicit learning abilities, facilitated the effect of implicit feedback on gender agreement. DOM depends on the semantic notion of animacy, which may contribute to making form-meaning mapping more transparent (see e.g., Williams, 1999). Such a meaningful relationship between abstract elements could have made this target structure more salient for those learners with higher explicit learning ability letting them rely on explicit cues. However, grammatical agreement is a formal, noninterpretable feature that requires associating nonmeaningful elements that co-occur. Learners with higher implicit learning ability were better at picking up these elements in the implicit feedback in an unconscious, nonreflective way.

The results regarding the role of cognitive abilities in explicit feedback are similar to those reported in Granena and Yilmaz's (2018) research synthesis, which showed that explicit cognitive abilities were positively related to learners' gains under explicit conditions, but implicit cognitive abilities were not. The results regarding the role of cognitive abilities in implicit feedback, however, diverge from Granena and Yilmaz (2018). In Granena and Yilmaz (2018), explicit cognitive abilities facilitated learning under implicit conditions, whereas implicit cognitive abilities inhibited learning. The current study, by contrast, found a facilitative role for implicit learning ability but no role for explicit learning. The discrepancy between Granena and Yilmaz (2018) and the current study regarding the role of implicit cognitive abilities in implicit instruction

https://doi.org/10.1017/S0272263120000601 Published online by Cambridge University Press

may be related to the fact that Granena and Yilmaz (2018) included both feedback and nonfeedback studies. In addition, Granena and Yilmaz (2018) warned that the trustworthiness of their results about the role of implicit cognitive abilities might have been low because only two of the synthesized studies included an implicit cognitive ability.

Regarding the three components that emerged from the cognitive measures, we labeled two of them as being implicit, implicit learning and implicit memory. However, only the one referring to implicit learning had an effect on learning outcomes. As discussed in Granena (2019), implicit learning mostly concerns the encoding of input, while implicit memory mostly concerns the retention and retrieval of information. Some individuals are better at spreading activation in implicit memory and, as a result, at making use of, or retrieving, information more efficiently (Woltz, 2003). In the context of the instructional intervention investigated, which targeted structures that learners had not fully acquired yet, one might expect a learning ability to play a more relevant role than a memory ability. In fact, in Granena (2019), the ability that showed a relationship with fluency scores in a picture description task used as a general proficiency measure was implicit memory. In such a context, which aimed at targeting general L2 speaking proficiency, the ability to recall words quickly and automatically (i.e., implicit memory ability) might have been more relevant than the ability to learn implicitly.

CONCLUSION

This ATI study investigated the extent to which cognitive abilities for implicit and explicit learning are differentially related to the acquisition of Spanish noun-adjective gender agreement and DOM under implicit and explicit feedback conditions. We found that learning gains in the implicit feedback group were related to implicit, but not explicit, learning ability, and that learning gains in the explicit feedback group were related to explicit, but not implicit, learning ability.

It is important to acknowledge that some of the specific methodological features of this study, including the choice of linguistic targets, feedback types, and language and cognitive measures, might limit the generalizability of our study. An important limitation of the study is the fact that it did not include any direct evidence of how participants processed the L2 under the feedback conditions. This limitation prevents us from making any claims regarding the nature of learning under different feedback conditions. Also, our operationalization of implicit feedback, that is, partial recasts, might have made implicit learning less likely because partial recasts have been considered to be relatively more salient than other types of recasts (Loewen & Philp, 2006). In addition, the following methodological improvements can be introduced to the designs of future studies to increase the trustworthiness of the findings: (a) feedback amount different feedback groups receive can be held constant or fewer opportunities for feedback can be given to better balance groups in feedback amount; and (b) more stringent participant selection criteria can be used to exclude learners who spoke a third language displaying the same linguistic phenomena as the target language.

Despite these limitations, our findings are consistent with a double dissociation between instructional condition and cognitive ability and have relevant theoretical and practical implications. Theoretically, it informs about the nature of the learning processes at work during the two instructional interventions and demonstrates that these processes

https://doi.org/10.1017/S0272263120000601 Published online by Cambridge University Press

were qualitatively different. Practically, it shows that differentiated instruction matched to learners' cognitive strengths may be more effective, at least with some L2 structures. ATI research, however, is a complex arena, and support for the double dissociation in this study was only partial. Therefore, more ATI studies are needed to clarify the extent to which the explicitness, or implicitness, of the instructional condition determines the explicitness, or implicitness, of the cognitive ability learners draw on when they learn under that condition.

SUPPLEMENTARY MATERIALS

To view supplementary material for this article, please visit http://dx.doi.org/10.1017/S0272263120000601.

NOTES

[1]Some researchers (e.g., Loewen & Philp, 2006) argue that the salience of recasts can be enhanced when they are provided with prosodic stress, in declarative intonation, or with other feedback moves such as an elicitation.

[2]We used a cloze test to measure proficiency because there is evidence showing that cloze tests correlate strongly with standardized measures of proficiency (Bachman, 1985). We also measured proficiency through an oral picture description task to triangulate the results of the cloze test with the results of a test that is more similar to the treatment task in task conditions and response modality.

[3]Hair et al. (2006) argued that reliability may be as low as 0.60 or 0.50 and still be acceptable in exploratory studies in the social sciences.

[4]Eight unique items were carried from the treatment to each version of each test type.

[5]The term "marked" is used to refer to forms that are nonbasic or less natural.

[6]As argued in Granena (2019), a possible explanation for the negative loading of the Letter span test could be the different nature of the cognitive processes involved in short-term memory and implicit learning. Unlike implicit learning, short-term memory engages attention-driven memory processes. The inverse relationship found could be interpreted as a competition of implicit and explicit cognitive processes in implicit learning such that restricting the contribution of the explicit processes involved in short-term memory results in an enhancement of implicit learning. The fact that the same relationship was not found in the case of explicit learning could indicate that explicit cognitive processes do not require restricting the contribution of implicit processes to operate.

[7]Following Cohen (1988), the strength of a linear relationship can be weak ($0 < r < 0.20$), moderately weak ($0.20 < r < 0.40$), moderate ($0.41 < r < 0.60$), moderately strong ($0.61 < r < 0.80$), and strong ($0.81 < r < 1.0$).

REFERENCES

Alarcón, I. (2010). Gender assignment and agreement in L2 Spanish: The effects of morphological marking, animacy and gender. *Studies in Hispanic and Lusophone Linguistics*, *3*, 267–299.

Anderson, J. R. (1987). Skill acquisition: Compilation of weak-method problem solutions. *Psychological Review*, *94*, 192–210.

Arroyo, D., & Yilmaz, Y. (2017). The role of language analytic ability in the effectiveness of different feedback timing conditions. In L. Gurzynski-Weiss (Ed.), *Expanding individual difference research in the interaction approach: Investigating learners, instructors, and other interlocutors* (pp. 71–97). John Benjamins.

Bachman, L. F. (1985). Performance on cloze tests with fixed-ratio and rational deletions. *TESOL Quarterly*, *19*, 535–555.

Baddeley, A. D., & Hitch, G. (1974). Working memory. In G. H. Bower (Ed.), *The psychology of learning and motivation: Advances in research and theory* (Vol. *8*, pp. 47–89). Academic Press.

Buffington, J., & Morgan-Short, K. (2019). Declarative and procedural memory as individual differences in second language aptitude. In Z. Wen, P. Skehan, A. Biedroń, S. Li, & R. Sparks (Eds.), *Language aptitude: Advancing theory, testing, research and practice* (pp. 215–237). Routledge.

https://doi.org/10.1017/S0272263120000601 Published online by Cambridge University Press

Carroll, J. B. (1962). The prediction of success in intensive foreign language training. In R. Glaser (Ed.), *Training research and education* (pp. 87–136). University of Pittsburgh Press.

Carroll, J. B. (1966). *A parametric study of language training in the Peace Corps. Final report.* Harvard University.

Carroll, J. B., & Sapon, S. (1959). *Modern Language Aptitude Test: Form A*. Psychological Corporation.

Carroll, S. (2001). *Input and evidence: The raw material of second language acquisition*. John Benjamins.

Carroll, S., & Swain, M. (1993). Explicit and implicit negative feedback: An empirical study of the learning of linguistic generalizations. *Studies in Second Language Acquisition*, *15*, 357–386.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Lawrence Erlbaum.

Conway, C. M., & Christiansen, M. H. (2005). Modality-constrained statistical learning of tactile, visual, and auditory sequences. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *31*, 24–39.

Cronbach, L., & Snow, R. (1977). *Aptitudes and instructional methods: A handbook for research on interactions*. Irvington.

Dean, A., & Voss, D. (1999). *Design and analysis of experiments*. Springer-Verlag. doi:10.1007/b97673.

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

DeKeyser, R. M., & Koeth, J. (2011). Cognitive aptitudes for second language learning. In E. Hinkel (Ed.), *Handbook of research in second language teaching and learning* (Vol. *2*, pp. 395–406). Routledge.

Doughty, C. J. (2013). Optimizing post-critical-period language learning. In G. Granena & M. H. Long (Eds.), *Sensitive periods, language aptitude, and ultimate L2 attainment* (pp. 153–174). Wiley-Blackwell.

Doughty, C. J. (2014). Assessing aptitude. In A. Kunnan (Ed.), *The companion to language assessment* (pp. 25–46). Wiley-Blackwell.

Doughty, C. J. (2019). Cognitive language aptitude. *Language Learning*, *69*(Suppl. 1), 101–126.

Ellis, R. (1991). *Instructed second language acquisition: Learning in the classroom*. Wiley-Blackwell.

Ellis, R. (2005). Measuring implicit and explicit knowledge of second language: A psychometric study. *Studies in Second Language Acquisition*, *27*, 141–172.

Ellis, R. (2007). The differential effects of corrective feedback on two grammatical structures. In A. Mackey (Ed.), *Conversational interaction in second language acquisition* (pp. 339–360). Oxford University Press.

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.

Engle, R. W., Tuholski, S. W., Laughlin, J. E., & Conway, A. R. A. (1999). Working memory, short-term memory, and general fluid intelligence: A latent-variable approach. *Journal of Experimental Psychology: General*, *128*, 309–331.

Fernández-García, M. (1999). Patterns of gender agreement in the speech of second language learners. In J. Gutierrez-Rexach & F. Martinez-Gil (Eds.), *Advances in Hispanic linguistics: Papers from the 2nd Spanish linguistics symposium* (Vol. *1*, pp. 3–15). Cascadilla.

Finneman, M. D. (1992). Learning agreement in the noun phrase: The strategies of three first-year Spanish students. *International Review of Applied Linguistics in Language Teaching*, *30*, 121–137.

Gass, S. M., & Mackey, A. (2006). Input, interaction and output: An overview. *AILA Review*, *19*, 3–17.

Gebauer, G. F., & Mackintosh, N. J. (2007). Psychometric intelligence dissociates implicit and explicit learning. *Journal of Experimental Psychology*, *33*, 34–54.

George, D., & Mallery, M. (2010). *SPSS for windows step by step: A simple guide and reference, 17.0 update* (10th ed.). Pearson.

Granena, G. (2013). Cognitive aptitudes for second language learning and the LLAMA Language Aptitude Test. In G. Granena & M. H. Long (Eds.), *Sensitive periods, language aptitude, and ultimate L2 attainment* (pp. 105–129). John Benjamins.

Granena, G. (2014). Language aptitude and long-term achievement in early childhood L2 learners. *Applied Linguistics*, *35*, 483–503.

Granena, G. (2019). Cognitive aptitudes and L2 speaking proficiency: Links between LLAMA and Hi-LAB. *Studies in Second Language Acquisition*, *41*, 313–336.

Granena, G., & Yilmaz, Y. (2018). Aptitude-Treatment Interaction in L2 learning: A research synthesis. *Studies in English Education*, *23*, 803–830.

Granena, G., & Yilmaz, Y. (2019). Corrective feedback and the role of implicit sequence learning ability in L2 online performance. *Language Learning*, *69*, 127–156.

https://doi.org/10.1017/S0272263120000601 Published online by Cambridge University Press

Goo, J. (2012). Corrective feedback and working memory capacity in interaction-driven L2 learning. *Studies in Second Language Acquisition*, *34*, 445–474.

Goo, J. (2016). Corrective feedback and working memory capacity: A replication. In G. Granena, D. O. Jackson, & Y. Yilmaz (Eds.), *Cognitive individual differences in second language processing and acquisition* (pp. 279–302). John Benjamins.

Goo, J., Granena, G., Yilmaz, Y., & Novella, M. (2015). Implicit and explicit instruction in L2 learning. In P. Rebuschat (Ed.), *Implicit and explicit learning of languages* (pp. 443–482). John Benjamins.

Guijarro-Fuentes, P. (2012). The acquisition of interpretable features in L2 Spanish: Personal a. *Bilingualism: Language and Cognition*, *15*, 701–720.

Hair, J. F., Black, W. C., Babin, B. J., Anderson, R. E., & Tatham, R. L. (2006). *Multivariate data analysis* (6th ed.). Pearson, Prentice Hall.

Hamrick, P. (2015). Declarative and procedural memory abilities as individual differences in incidental language learning. *Learning and Individual Differences*, *44*, 9–15.

Hayes, A. F. (2017). *Introduction to mediation, moderation, and conditional process analysis: A regression-based approach*. Guilford Press.

Hayes, A. F. (n.d.). *PROCESS* (Computer Software). Retrieved from https://www.processmacro.org/download.html.

Hulstijn, J. H. (2015). Explaining phenomena of first and second language acquisition with the constructs of implicit and explicit learning: The virtues and pitfalls of a two-system view. In P. Rebuschat (Ed.), *Implicit and explicit learning of languages* (pp. 25–46). John Benjamins.

Kalra, P., Gabrieli, J. D. E., & Finn, A. S. (2019). Evidence of stable individual differences in implicit learning. *Coqgnition*, *190*, 199–211.

Kane, M. J., & Engle, R. W. (2002). The role of prefrontal cortex in working memory capacity, executive attention, and general fluid intelligence: An individual difference perspective. *Psychonomic Bulletin and Review*, *9*, 637–667.

Kaufman, S. B., DeYoung, C. G., Gray, J. R., Jimenez, L., Brown, J., & Mackintosh, N. (2010). Implicit learning as an ability. *Cognition*, *116*, 321–340.

Krashen, S. (1981). *Second language acquisition and second language learning*. Pergamon Press.

Kyllonen, P. C. (1996). Is working memory capacity Spearman's *g*? In I. Dennis & P. Tapsfield (Eds.), *Human abilities: Their nature and measurement* (pp. 49–75). Lawrence Erlbaum Associates, Inc.

Kyllonen, P. C., & Christal, R. E. (1990). Reasoning ability is (little more than) working-memory capacity? *Intelligence*, *14*, 389–433.

Leeman, J. (2003). Recasts and second language development. *Studies in Second Language Acquisition*, *25*, 37–63.

Li, S. (2010). The effectiveness of corrective feedback in SLA: A meta-analysis. *Language Learning*, *60*, 309–365.

Li, S. (2013). The interactions between the effects of implicit and explicit feedback and individual differences in language analytic ability and working memory. *The Modern Language Journal*, *97*, 634–654.

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

Long, M. H. (1996). The role of the linguistic environment in second language acquisition. In W. Ritchie, & T.K. Bhatia (Eds.), *Handbook of second language acquisition* (pp. 413–468). Academic Press.

Long, M. H. (2007). *Problems in SLA*. Lawrence Erlbaum.

Long, M. H., & Doughty, C. (2009). *The handbook of language teaching*. Wiley-Blackwell.

Loewen, S., & Philp, J. (2006). Recasts in the adult L2 classroom: Characteristics, explicitness and effectiveness. *The Modern Language Journal*, *90*, 536–556.

Lyster, R., & Saito, K. (2010). Oral feedback in classroom SLA: A meta-analysis. *Studies in Second Language Acquisition*, *32*, 265–302.

Mackey, A., & Goo, J. (2007). Interaction research in SLA: A meta-analysis and research synthesis. In A. Mackey (Ed.), *Conversational interaction in SLA: A collection of empirical studies* (pp. 408–452). Oxford University Press.

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

Montrul, S., & Gürel, A. (2015). The acquisition of Differential Object Marking in Spanish by Turkish speakers. In T. Judy & S. Perpiñán (Eds.), *The acquisition of Spanish by speakers of less commonly studies languages* (pp. 281–308). John Benjamins.

https://doi.org/10.1017/S0272263120000601 Published online by Cambridge University Press

Morgan-Short, K., Faretta-Stutenberg, M., & Bartlett-Hsu, L. (2015). Contribution of event-related potential research into explicit and implicit second language acquisition. In P. Rebuschat (Ed.), *Investigating implicit and explicit language learning* (pp. 349–386). John Benjamins.

Morgan-Short, K., Faretta-Stutenberg, M., Brill-Schuetz, K., Carpenter, H., & Wong, P. C. M. (2014). Declarative and procedural memory as individual differences in second language acquisition. *Bilingualism: Language and Cognition*, *17*, 56–72.

Nassaji, H. (2015). *The interactional feedback dimension in instructed second language learning: Linking theory, research and practice*. Bloomsbury.

Norris, J. M., & Ortega, L. (2000). Effectiveness of L2 instruction: A research synthesis and quantitative meta-analysis. *Language Learning*, *50*, 417–528.

Parry, T. S., & Child, J. R. (1990). Preliminary investigation of the relationship between VORD, MLAT, and language proficiency. In T. S. Parry & C. W. Stansfield (Eds.), *Language aptitude reconsidered* (pp. 30–66). Prentice-Hall Regents.

Reber, A. S. (1993). *Implicit learning and tacit knowledge: An essay on the cognitive unconscious*. Oxford University Press.

Revesz, A. (2012). Working memory and the observed effectiveness of recasts on different l2 outcome measures. *Language Learning*, *62*, 93–132.

Seger, C. A. (1994). Implicit learning. *Psychological Bulletin*, *115*, 163–196.

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). Oxford University Press.

Sheen, Y., & Ellis, R. (2011). Corrective feedback in L2 teaching. In E. Hinkel (Ed.), *Handbook of research in second language teaching and learning* (pp. 593–610). Routledge.

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

Tagarelli, K. M., Ruiz, S., Vega, J. L. M., & Rebuschat, P. (2016). Variability in second language learning: The roles of individual differences, learning conditions, and linguistic complexity. *Studies in Second Language Acquisition*, *38*, 293–316.

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* (pp. 144–171). Oxford University Press.

Williams, J. (1999). Memory, attention, and inductive learning. *Studies in Second Language Acquisition*, *21*, 1–48.

Woltz, D. J. (2003). Implicit cognitive processes as aptitudes for learning. *Educational Psychologist*, *38*, 95–104.

Yilmaz, Y. (2012). The relative effects of explicit correction and recasts on two target structures via two communication modes. *Language Learning*, *62*, 1134–1169.

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

Yilmaz, Y. (2013b). The relative effectiveness of mixed, explicit and implicit feedback. *System*, *41*, 691–705.

Yilmaz, Y. (2016). The linguistic environment, interaction and negative feedback. *Brill Research Perspectives in Multilingualism and Second Language Acquisition*, *1*, 45–86.

Yilmaz, Y., & Granena, G. (2016). The role of cognitive aptitudes for explicit language learning in the relative effects of explicit and implicit feedback. *Bilingualism: Language and Cognition*, *19*, 147–161.

Yilmaz, Y., & Granena, G. (2019). Cognitive individual differences as predictors of improvement and awareness under implicit and explicit feedback conditions. *Modern Language Journal*, *103*, 686–702.

Yilmaz, Y., & Koylu, Y. (2016). The interaction between feedback exposure condition and phonetic coding ability. In G. Granena, D. O. Jackson, & Y. Yilmaz (Eds.), *Cognitive individual differences in second language processing and acquisition* (pp. 303–326). John Benjamins.

https://doi.org/10.1017/S0272263120000601 Published online by Cambridge University Press