Variability in second language learning: The roles of individual differences, learning conditions, and linguistic complexity

Kaitlyn M. Tagarelli[1]
Simón Ruiz-Hernández [2]
José Luis Moreno Vega[3]
Patrick Rebuschat[3]

[1] Department of Psychology and Neuroscience, Dalhousie University
[2] LEAD Graduate School, Eberhard Karls Universität Tübingen
[3] Department of Linguistics and English Language, Lancaster University

Correspondence concerning this article should be addressed to Kaitlyn M. Tagarelli, Dept. of Psychology & Neuroscience, Life Sciences Center, 1355 Oxford St., Dalhousie University, Halifax, NS B3H 4R2.
Email: kaitlyn.tagarelli@dal.ca

Abstract

Second language learning outcomes are highly variable, due to a variety of factors, including individual differences, exposure conditions, and linguistic complexity. However, exactly how these factors interact to influence language learning is unknown. This paper examines the relationship between these three variables in language learners.

Native English speakers were exposed to an artificial language containing three sentence patterns of varying linguistic complexity. They were randomly assigned to two groups – incidental and instructed – designed to promote the acquisition of implicit and explicit knowledge, respectively. Learning was assessed with a grammaticality judgment task, while subjective measures of awareness were used to measure whether exposure had resulted in implicit or explicit knowledge. Participants also completed cognitive tests.

Awareness measures demonstrated that learners in the incidental group relied more on implicit knowledge, whereas learners in the instructed group relied more on explicit knowledge. Overall, exposure condition was the most significant predictor of performance on the grammaticality judgment task, with learners in the instructed group outperforming those in the incidental group. Performance on a procedural learning task accounted for additional variance. When outcomes were analysed according to linguistic complexity, exposure condition was the most significant predictor for two syntactic patterns, but it was not a predictor for the most complex sentence group; instead, procedural learning ability was.

Introduction

Second language (L2) learning outcomes vary considerably. The source of this variation appears to be multifaceted, stemming not only from individual differences (IDs) in cognitive abilities (e.g., Dörnyei, 2006), but also from differences in affective factors like personality and motivation (e.g., Grey, Williams, & Rebuschat, 2015; Masgoret & Gardner, 2003), language exposure conditions (e.g., Godfroid, this issue; Granena, 2013; 2015; Morgan-Short, Sanz, Steinhauer, & Ullman, 2010; Rebuschat, 2008; Tagarelli, Borges Mota, & Rebuschat, 2015), and linguistic complexity (e.g., DeKeyser, 2005; Hawkins, 2004). While examining these factors in isolation can inform the nature of L2 acquisition, we can learn even more by investigating how multiple factors interact to mediate learning processes and outcomes. This paper aims to contribute to our understanding of the sources of variation in L2 success by focusing on three major factors that have been shown to influence it – IDs in cognitive abilities, exposure conditions, and linguistic complexity – and the interactions between them.

Individual Differences

According to Skehan (1998), IDs in aptitude could have major relevance for language development, inasmuch as this development requires a certain capacity to process and restructure input. Indeed, IDs such as aptitude, working memory (WM), personality, attitude, motivation, learning styles, and learning strategies, have been shown to influence L2 processing and achievement (Dörnyei, 2006; Juffs & Harrington, 2011).

Individual differences and exposure conditions. While IDs may predict L2 success, the ways in which cognitive abilities interact with instructional methods cannot be ignored. In L2 research, language exposure is often manipulated on a continuum from *explicit* conditions, in which learners are provided with metalinguistic information (e.g. pedagogical rules) or instructed to look for rules or patterns, to *implicit* conditions, in which such information or instruction is not provided (DeKeyser, 1995). It is important to note, however, that while these conditions are often designed to respectively promote the development of explicit (conscious) or implicit (unconscious) knowledge, it is not necessarily the case that such knowledge is acquired. In fact, several studies have shown that implicit conditions can lead to the development of explicit knowledge, and vice versa (Godfroid, this issue, Granena 2013; Grey, Williams, & Rebuschat, 2014; Rebuschat, 2008; Rebuschat & Williams, 2012; Rebuschat, Hamrick, Riestenberg, Sachs, & Ziegler 2015; Tagarelli et al., 2015). There has been some disagreement as to whether more structured, explicit training conditions should amplify (Krashen, 1981) or level-out (Skehan, 1989) the effects of IDs on learning, but there is emerging evidence that individual learners might not benefit equally from the same method of instruction (de Graaff, 1997; Erlam, 2005; Robinson, 2002, 2005; Tagarelli et al., 2015).

This apparent interaction between IDs and exposure conditions could stem from differences between explicit and implicit processes. Traditionally, cognitive psychology suggests that implicit processes should be less susceptible to IDs than explicit processes (Reber, 1993; Reber et al., 1991). However, more recent work has demonstrated IDs in implicit learning abilities (Granena, 2013; Kaufman, DeYoung, Gray, Jiménez, Brown, & Mackintosh, 2010), and several studies have found relationships between IDs and learning outcomes in implicit exposure conditions (Grey, Williams, & Rebuschat, 2015; Hamrick, in press; Morgan-Short, Faretta-Stutenberg, Brill-Schuetz, Carpenter, & Wong, 2014; Robinson, 2002; Williams, 2003; Williams & Lovatt, 2003). These inconsistencies likely arise from the fact that certain cognitive abilities may be associated with implicit or explicit processes, and therefore have differential effects on learning under these conditions. Two cognitive domains that are particularly relevant to this issue are WM and procedural memory.

Working memory. WM is the system of temporary storage and manipulation of information during complex cognitive activities such as language comprehension and learning (Baddeley, 2010). There are several models of WM, but here we assume the multi-component model proposed by Baddeley and colleagues (Baddeley, 2010; Baddeley & Hitch, 1974). This consists of domain-specific storage components for auditory and visuospatial information, an episodic buffer that links these storage components to long-term memory, and the central executive, a domain-general, limited-capacity construct that is responsible for the allocation of attention.

WM is widely thought to play an important role in L2 learning (Hummel, 2009; Juffs & Harrington, 2011; Linck, Osthus, Koeth, & Bunting, 2014; Martin & N. Ellis, 2012; Williams, 2012). According to Baars (1988), WM allows explicit deduction, hypothesis formation, analogical reasoning, prioritization, control, and decision making. Roehr (2008) suggests a possible link between WM and explicit exposure conditions, particularly when L2 learners need to retain metalinguistic information while comprehending and producing language (see also Clark & Squire, 1988). WM has been shown to impact L2 classroom learning, which is often associated with explicit processes (Linck & Weiss, 2011). Some research suggests that WM is strongly related to controlled, explicit language processing, but also plays a moderate a role in identifying tasks that can be delegated to more implicit, automatic processing, including less monitored speaking and listening (Erçetin & Alptekin, 2013). Overall, it seems that WM may underlie both implicit and explicit language processes, but is more closely tied to explicit ones.

Procedural memory. The procedural memory system is one of several long-term memory systems in the brain (Squire & Knowlton, 2000). It is likely involved in the acquisition, storage, and use of information that requires coordination of actions or sequences, like driving a car. In L1 and highly proficient L2, procedural memory is thought to underlie the learning, storage, and processing of grammar (Ullman, 2015). Because it relies on an implicit memory system, procedural learning ability may reliably predict L2 learning in implicit conditions, at least after learners have high levels of proficiency and/or exposure to the language (Morgan-Short et al., 2014).

Linguistic Complexity

The relationship between IDs and training conditions may be mediated by a third factor: the complexity of the structures being learned. This paper focuses on linguistic complexity, and more specifically *structure complexity*, or “the individual linguistic items, structures, or rules that make up the learner’s L2 system” (Bulté & Housen, 2012, p. 25). Early research suggested that the complexity of linguistic structures interacts with type of instruction. For instance, explicit instruction may be beneficial when learners are exposed to simple patterns (e.g., Krashen, 1994), whereas patterns that are sufficiently complex are more likely to be acquired by an implicit learning mechanism (e.g., Reber, 1993). Despite some evidence to support these predictions, the findings are inconsistent, with some suggesting that explicit conditions are more conducive to learning complex rules (see Housen, 2014; Spada & Tomita, 2010). One plausible reason for this is the lack of a consensus in operationalizing rule complexity (Bulté & Housen, 2012). Another possibility is that the learnability of simple and complex rules in implicit and explicit conditions may be mediated by a third variable, IDs. The current study systematically manipulates these three factors with the goal of understanding their individual and combined roles in L2 learning and processing.

The Current Study

In this study, participants were trained on an artificial language in either incidental or instructed exposure conditions. Importantly, this study did not assume that these learning contexts would lead to implicit and explicit learning or knowledge, respectively. The primary outcome measure in this study was performance on an aural untimed grammaticality judgment task (GJT), which is often categorized as an explicit knowledge measure (see R. Ellis, 2005). In order to obtain a more fine-grained measure of whether our exposure conditions resulted in conscious or unconscious knowledge, online and offline measures of awareness were included in the experimental design (Dienes & Scott, 2005; see Rebuschat, 2013, for review). Participants also completed cognitive tasks designed to assess WM and procedural learning ability.

Artificial Language. An artificial language is a model linguistic system comprised of a small, novel lexicon and a few grammatical rules that are consistent with natural language rules (e.g., de Graaff, 1997; DeKeyser, 1995; Yang & Givón, 1997). Like natural languages, they can be spoken and understood, but their small size and novelty allow them to be learned relatively quickly, and in highly controlled settings. Semi-artificial languages, in contrast, often combine lexical information from the participants' L1 and grammatical information from another language (e.g., Alanen, 1995; Grey, Williams, & Rebuschat, 2014; Hamrick, 2014; Leung & Williams, 2012; Rogers, Révész, & Rebuschat, in press; Williams, 2005). This facilitates learning even more, as participants do not need to learn new vocabulary.

The stimuli for this experiment were generated from a semi-artificial language consisting of an English lexicon and German syntax (see also Rebuschat, 2008; Rebuschat & Williams, 2012; Tagarelli et al., 2015). The linguistic focus was on three verb placement rules. These rules state that, depending on the type of clause (main vs. subordinate) and clause sequence (main-subordinate vs. subordinate-main), finite verbs have to be placed in either first, second or final position. Each rule (1-3) is associated with a specific syntactic pattern, as illustrated below. Example sentences are shown in Table 1.

(1) Simple: The finite verb is placed in second phrasal position of main clauses that are not preceded by a subordinate clause. (V2 pattern)

(2) Complex 1: The finite verb is placed in final position in all subordinate clauses. (V2-VF pattern)

(3) Complex 2: When a subordinate clause precedes a main clause, the finite verb is placed in first position in the main clause and final position in the subordinate clause. (VF-V1 pattern)

These three rules and associated patterns were designed to vary according to complexity. One way to assess structural complexity is through T-units, which are defined as "one main clause with all subordinate clauses attached to it" (Hunt, 1965, p. 20). Linguistic complexity can thus be defined as the number of clauses per T-unit (e.g., Larsen-Freeman, 1983). Based on this definition, the V2 pattern is the least complex construction, as it only contains one clause. The V2-VF and the VF-V1 patterns are more complex because they each contain two clauses. This assessment of complexity is also upheld in terms of cognitive complexity, or difficulty. Single-clause sentences are easier to process than those with multiple clauses (Bygate, 1999; Lord, 2002), and both L1 and L2 development are characterized by an increased production of sentences with subordinate clauses (Brandt, Diessel & Tomasello, 2008; Scott & Tucker, 1974; Gaies, 1976; Larsen-Freeman, 1978), suggesting that clause length is related to difficulty. Moreover, while the V2-VF and VF-

V1 patterns are similar in terms of linguistic T-unit complexity, it could be argued that the VF-V1 pattern is cognitively more complex or difficult because the subordinate clause precedes the main clause, which may further increase the processing load (Jarvella & Herman, 1972).

This study aimed to investigate whether WM and procedural learning ability would predict L2 learning outcomes, and to determine whether these predictions would vary according to exposure condition and structural complexity. We set out to address three research questions:

(1) What is the relationship between exposure condition and syntactic complexity? In other words, does exposure condition (incidental vs. instructed) differentially affect the learnability of syntactic structure?

(2) What is the relationship between the cognitive factors under investigation (WM and procedural learning) and learning outcomes in incidental and instructed exposure conditions?

(3) Is there an interaction between cognitive abilities, exposure conditions, and syntactic complexity in predicting L2 outcomes?

Regarding RQ1, we expected overall performance to be best on the simplest structure (V2), followed by the complex structures (V2-VF and VF-V1). We might expect poorer performance on Complex 2 (VF-V1) relative to Complex 1 (V2-VF) sentences, since the Complex 2 structure may be more cognitively demanding, as outlined above. Considering exposure conditions, we hypothesized that learning of simple rules would be better in the instructed condition and learning of complex rules would be better in the incidental condition (Reber, 1993).

Regarding RQ2, we hypothesized that WM span would correlate with performance in the instructed condition, but not in the incidental condition, given previous research suggesting that it is more strongly linked to explicit language processes (Baars, 1988; Clark & Squire, 1998; Erçetin & Alptekin, 2013). We expected the opposite effect for implicit/procedural learning abilities (Morgan-Short et al., 2014).

Our third research question is largely exploratory, and thus we had no specific predictions regarding the three-way interaction between IDs, exposure conditions, and complexity.

Method

Participants

Fifty-one native speakers of English with no background in German (or any other V2 language) were randomly assigned to an incidental exposure condition (n = 25, 19 female, M_{age} = 19.30) or to an instructed exposure condition (n = 25, 20 female, M_{age} = 19.32). One additional participant was excluded for failing to follow task instructions. All participants were exposed to the same artificial language stimuli. However, participants in the incidental group were not informed of the linguistic target, nor that there would be a testing phase. In contrast, participants in the instructed group were explicitly taught the target rule system before being exposed to the language. Data was collected in the US and in the UK (see Section S1 in Supplementary Materials for detailed information about participant groups). Participants received either course credit or monetary compensation.

Materials & Procedure

Participants first completed a language learning session, followed by cognitive tests on a separate day. Each session lasted approximately one hour.

Session 1: Artificial language training and testing. The language session consisted of training and a GJT. The tasks and materials were exactly the same for both conditions, except that the instructed group was provided with rules before the actual training phase.

Training set. The training set consisted of 120 sentences, with 40 sentences for each syntactic pattern. For each pattern, half the sentences were semantically plausible (see Table 1) and the other half semantically implausible (e.g., "Rose abandoned in the evening her cats on planet Venus"). Plausible and implausible items were designed so that participants would have to process the entire auditory string before judging its plausibility.

Testing set. The testing set consisted of 60 new sentences, 30 grammatical and 30 ungrammatical. All sentences were semantically plausible. There were six ungrammatical sentence types that were matched to specific grammatical patterns for the purposes of *d*-prime analyses. The "Simple" V2 pattern was matched to *V3 and *V4, two ungrammatical sentence types that also only contained a single clause. The "Complex 1" V2-VF pattern was matched to the *VF pattern, which had a verb in final position in the main clause, and to the *V1-VF pattern, which had correct verb placement in the subordinate clause, but not in the main clause. Finally, the "Complex 2" VF-V1 pattern was matched to the *V1 pattern, which has a verb in initial position in the main clause, and to the *VF-V2 pattern, which again has a correct subordinate clause but incorrect main clause (see Table 1).

[INSERT TABLE 1 ABOUT HERE]

Training task. Participants in the *incidental* group were told that they were being tested on their ability to understand the meaning of sentences with scrambled word order, and they were not told that there would be a test on word order. For each trial, they were asked to listen to a sentence, judge the plausibility of the sentence, and then repeat each sentence after a delayed prompt (1500 ms). The focus of their task was therefore primarily on meaning. There were four practice trials and 120 randomly presented training sentences.

Participants in the *instructed* group were told that there was a complex rule system that determined the word order of the artificial language. They were then presented with each of the verb-placement rules and asked to write down two sentences that followed each rule. After this, they engaged in the same meaning-focused task as the incidental group (plausibility judgments, followed by sentence repetitions), but they were instructed to think about the verb-placement rules as they completed the task. The focus of their task was therefore on both form and meaning.

Grammaticality judgment task. After training, participants completed an untimed GJT. Before the test, participants in the incidental group were told (and participants in the instructed group were reminded) that the word order of the sentences presented during training followed a rule system, and that they should judge the new sentences based on the sentences they had heard in training. After four practice trials, all participants listened to the 60 test sentences, which were presented randomly. For each sentence, participants were asked to judge its grammaticality, indicate how confident they were in the accuracy of their judgment, and report the source of their judgment. These confidence ratings and source attributions served as subjective measures of awareness (Dienes & Scott, 2005; Rebuschat, 2013).

Participants indicated their levels of confidence by selecting one of four response options: *not at all confident*, *somewhat confident*, *quite confident*, and *extremely confident*. This was explained to the participants as a continuum of confidence from *not at all* to *extremely/100 percent*. There were also four response options for source attributions: *guess, intuition*, *memory,* and *rule* (see Section S.2 for definitions). All responses were entered on a serial response pad.

After completing the test phase, participants filled out a debriefing questionnaire, providing information on their levels of awareness regarding the rule system. The questionnaire also included biographical data, such as age, gender, occupation, major field of study, and language background.

Session 2: Cognitive tasks. During the second session, participants completed the RSpan and either the SRT or ASRT tasks.[1] Task order was randomized for each participant.

Reading Span Task. The Reading Span task (RSpan; Daneman & Carpenter, 1980) is a complex WM span task, meaning that it measures both the storage and processing components of WM. The RSpan task used in this study was adapted from Unsworth, Heitz, Schrock, and Engle (2005) and presented in Microsoft PowerPoint. In this task, participants saw a sentence appear on the screen. They were asked to read the sentence out loud, judge its plausibility, and then read a capitalized letter out loud (speaking out loud was intended to prevent subvocal rehearsal of previous letters in a set). Between two and five sentences were presented in a set, and set sizes were intermixed to prevent strategies (e.g., creating mnemonics of specific lengths based on set size). At the end of a set, participants were prompted to write down all letters in the sequence in which they had been presented, leaving a blank space for letters they could not remember. There were three practice sets and twelve test sets (three per set size).

Following Unsworth et al. (2005), the RSpan score was calculated as the sum of all set sizes that were perfectly recalled. The highest possible score was 42. All participants responded correctly to at least 80% of plausibility judgments, suggesting that they were both storing letters and processing sentences.

(Alternating) Serial Reaction Time. Procedural learning ability was measured by a Serial Reaction Time (SRT) and Alternating Serial Reaction Time (ASRT) tasks. Prior research suggests that SRT tasks provide the best measure of implicit/procedural learning ability (Kaufman et al., 2010; Shanks, 2005; Howard & Howard, 1997). Explicit learning is minimized in these tasks because participants are not given instructions that direct their attention to underlying rules, and there is no separate testing phase, so participants are never informed that rules exist (e.g., Kaufman et al., 2010).

In both tasks, participants saw four squares or circles arranged horizontally on a screen. On each trial, an object appeared in one of the locations, and the participant pressed a button corresponding to the location of the object on a serial response pad. Once the participant pressed the button, the object disappeared and then reappeared in a new position. The SRT (Nissen & Bullemer, 1987) consisted of six blocks with 60 trials per block. In the first and sixth blocks, the location of the object was random. In the other blocks, the location followed a 10-trial sequenced pattern. In this task, decreased reaction time over the course of the sequence blocks, followed by an increase in the final block, is taken as evidence that participants have learned the sequence, and are not simply getting faster at the button-pressing task. The ID score is the difference in average reaction time between the fifth and sixth blocks. A larger difference indicates a greater learning effect. The SRT was run on E-Prime 2.0 on a PC computer.

In the ASRT, random and sequence trials are interspersed within blocks, rather than allocated to separate blocks; every *other* trial follows a pattern, and the alternating trials are random (see Howard & Howard, 1997). The task consisted of 20 blocks of 85 trials each. It was run using E-Prime Version 1.2 on a PC computer. Following Kaufman et al. (2010), participants were awarded one point for every block in which a learning effect was observed (i.e., faster reaction time for sequence than random trials). The total score for the participant is the sum of scores across the last 18 blocks, for a maximum score of 18.

ASRT has several advantages over the SRT, which are addressed in detail by Howard and Howard (1997). We were unable to run the ASRT in the UK for technical reasons. However, while perhaps not ideal, SRT tasks have been successfully used to measure implicit learning (e.g., Kaufman et al., 2010) and serve as an acceptable alternative.

Results

This study aimed to examine the interaction between exposure conditions, syntactic complexity, and individual variation in cognitive abilities. We will address each of these in turn, but first, we need to establish (i) whether exposure to the artificial language resulted in learning, (ii) whether the knowledge acquired was implicit and/or explicit, and (iii) whether these outcomes varied according to exposure condition (incidental vs. instructed). This will allow us to place the results of this study within the framework of SLA broadly, and more specifically in the discussion of implicit and explicit processes in L2 acquisition.

Performance on the grammaticality judgment task

Learning was assessed by accuracy and d-prime (*d'*) scores on the GJT. Unlike raw accuracy measures, *d'* takes response bias into account in tasks with binary response options. It is calculated based on the hit and false alarm rates, and provides an index of how sensitive that particular subject is to the difference between grammatical and ungrammatical items.

Accuracy and *d'* scores for both groups are shown in Table 2. *D'* is plotted in Figure 1. Although *d'* scores were generally low, both the incidental group, $t(24) = 2.951$, $p = .007$, and the instructed group, $t(24) = 4.918$, $p < .001$, performed significantly above chance, indicating a clear learning effect in both exposure conditions. An independent samples *t*-test showed that the instructed group significantly outperformed the incidental group in terms of overall *d'* scores, $t(48) = 3.07$, $p = .004$.

[INSERT FIGURE 1 ABOUT HERE]

[INSERT TABLE 2 ABOUT HERE]

Subjective measures of awareness

For reasons of space, only results from source attributions are discussed here (see Section S3 for tables and additional information about subjective measures of awareness). Table S3.1 summarizes overall accuracy and proportions across the four source attribution categories for both groups. Participants in the incidental group only performed significantly above chance when basing grammaticality judgments on intuition. When attributing grammaticality decisions to memory or rule knowledge, their performance was indistinguishable from chance. Interestingly, the opposite pattern was observed in the instructed group. Here, participants only performed significantly above chance when basing grammaticality judgments on rule knowledge. A chi-square analysis revealed a highly significant association between training condition and source attributions, χ^2 (1) = 143.51, $p < .001$, in that the incidental group relied more on the implicit categories (guess and intuition) than the instructed group, and the instructed group more on the explicit categories (memory and rule knowledge) than the incidental group.

Exposure conditions and linguistic complexity

To answer RQ1, we looked at performance in the two exposure conditions across the three syntactic patterns (see Table 2 and Figure 1). Again, *d'* scores are relatively low, with the maximum score being 1.00 (corresponding to about 75% accuracy) for Simple sentences in the Instructed condition. Nevertheless, there are still evident differences in accuracy according to syntactic complexity. Participants in both conditions performed above chance on the Simple structures, and those in the instructed condition also performed above chance on both types of complex sentences. The *d'* scores for participants in the incidental group were close to zero for the Complex 1 sentences but approaching significance for Complex 2 sentences. Both groups show a trend of performing best on Simple structures, followed by Complex 2 and then Complex 1 sentences. The difference in *d'* scores between Simple and Complex 1 sentences was significant for both groups ($p < .03$).

A 2 x 3 mixed-effects repeated-measures ANOVA with Condition (Incidental vs. Instructed) and Complexity (Simple, Complex 1, Complex 2) showed that there were main effects of Condition, $F(1, 48) = 8.592$, $p = .005$, partial $\eta^2 = .152$, observed power = .819, and of Complexity, $F(2, 96) = 6.310$, $p = .003$, partial $\eta^2 = .116$, observed power = .889. There was no Complexity X Condition interaction, $F < 1$, which corroborates the similar trends in both groups. Bonferroni corrected post-hoc comparisons showed significantly better *d'* scores on Simple sentences than Complex 1 sentences, $p < .001$, but no other significant differences between sentence types. Learners in the instructed condition performed significantly better on Simple structures than those in the incidental condition, $p = .007$, and marginally better on Complex 2 sentences, $p = .039$ (corrected $\alpha = .017$). There were no differences between conditions on the Complex 2 sentences, but note the very high variance in both conditions.

Cognitive Measures

To answer RQ2, we examined performance on the cognitive tests, as well as the relationship between these variables and learning outcomes in the two exposure conditions. Scores from the ASRT and SRT were converted into z-scores (SRTz) and combined into one variable. Average performance on the RSpan and (A)SRT (and results of between-groups comparisons) are shown in Table S4.1.

Pearson's correlations were performed separately by group for each of the cognitive measures and *d'* scores on the GJT. There was a strong significant negative correlation between the SRTz and *d'* scores for the incidental group, $r = -.586$, $p = .003$, and no correlation between

RSpan and *d'* scores. There were no correlations between the cognitive variables and overall GJT performance for the instructed group (see Table S4.2).

Condition, Linguistic Complexity, and Cognitive Abilities

RQ3 addressed the three-way relationship between exposure conditions, linguistic complexity, and cognitive abilities in modulating learning outcomes. To examine this, we first ran correlations between performance on the ID tests and the GJT for each sentence type, separately for each exposure condition. This revealed that the correlation between the SRT and GJT in the incidental condition was mostly driven by the Complex 2 sentences, $r = -.543$, $p = .007$. There was also a medium-to-large positive correlation between the RSpan and *d'* scores in this condition, $r = .431$, $p = .035$. Notably, the RSpan and SRTz scores in the incidental condition were strongly negatively correlated, $r = -.501$, $p = .018$. These effect sizes are all rather large, though none withstand the correction for multiple comparisons, $\alpha = .003$. There were no significant correlations between RSpan or SRTz and *d'* scores for any sentence groups in the instructed condition.

To determine whether exposure condition, WM, or procedural learning ability predicted performance on the GJT for each sentence group, we entered these factors into a stepwise regression (see Table S5.1). Based on the results reported above and other research showing a relationship between the explicitness of training conditions and learning outcomes (Godfroid, this issue; Granena, 2013; 2015; Morgan-Short et al., 2010; Morgan-Short, Steinhauer, et al., 2012; Morgan-Short, Finger, et al., 2012; Rebuschat, 2008; Tagarelli et al., 2015), we entered exposure condition in the first step of the regression, followed by cognitive variables in the second step.

In the first models, exposure condition accounted for 14.3% and 9.9% of the variance for Simple and Complex 1 sentences, respectively, and was a significant predictor for both, $p < .05$. It only accounted for 4.1% of the variance for Complex 2 sentences, and was not a significant predictor, $p = .181$. The addition of the cognitive variables for Simple and Complex 1 accounted for slightly more variance, but neither was a significant predictor. For the Complex 2 sentences, the addition of the cognitive variables accounted for an additional 10% of the variance, and significantly improved the model. Specifically, performance on the SRT was a significant predictor, $p < .05$. All effect sizes were quite small.

Discussion

This study confirmed that adult learners can acquire L2 syntax under both instructed and incidental exposure conditions, while processing sentences for meaning, and without the benefit of feedback, thus replicating Rebuschat (2008, Expt. 3; Rebuschat & Williams, 2012, Expt. 2) and Tagarelli et al. (2015). We also demonstrated that the addition of metalinguistic rule presentation prior to meaning-oriented exposure results in a greater learning effect, consistent with previous studies (e.g., DeKeyser, 1995; de Graaff, 1997; Norris & Ortega, 2000; Spada & Tomita, 2010; Tagarelli et al., 2015). Our study also sheds light on the effect of different explicit exposure conditions on L2 acquisition. A comparison between the current study and Tagarelli et al. (2015) demonstrates that providing a brief metalinguistic treatment before meaning-oriented exposure resulted in similar gains to asking participants to consciously focus on form (discovering word-order rules) and disregard meaning (learning outcomes were statistically comparable, $t < 1$).

Our analysis of the subjective measures of awareness demonstrated that our exposure conditions resulted in the kind of knowledge they were designed to promote. Learners in the incidental condition relied more on the implicit categories (guess and intuition) and performed best when basing their judgments on intuition, and learners in the instructed condition relied more on the explicit categories (memory and rule) and performed best when basing their judgments on rule knowledge. In other words, the incidental group appears to have mainly acquired implicit knowledge, whereas the instructed group mainly acquired explicit knowledge. These results contrast with previous research showing that subjects often acquire both types of knowledge, with the relative amount varying according to learning context (e.g., Godfroid, this issue; Granena, 2013; Grey et al., 2014; Rebuschat, 2008; Rebuschat et al., 2015; Rogers et al., in press; Tagarelli et al., 2015).

Research Question 1: Exposure Conditions and Syntactic Complexity

We found no relationship between exposure condition and syntactic complexity. While the instructed group performed better overall, this was true for all sentence types, and the pattern of performance according to complexity was the same for both groups. Therefore, our hypothesis that linguistically simple rules would be easier to learn in the instructed condition and complex rules would be easier to learn in the incidental condition was not supported – simple rules were learned better in both conditions. This finding contradicts previous research suggesting that explicit instruction may be more beneficial for learning simple structures, given that these may be more available to conscious processing (DeKeyser, 1995; Reber, 1993), as well as other research showing an advantage of explicit instruction on the learning of complex structures (de Graaff, 1997; R. Ellis, 2007; Housen, Pierrard, & Van Daele, 2005). There is some evidence, however, that explicit instruction may result in similar outcomes for simple and complex structures (de Graaff, 1997). In addition, a significant difference between the Simple and Complex 1 sentences in the incidental condition contrast previous research suggesting that complexity may not be an important factor in determining learning outcomes in more implicit conditions (Krashen, 1994; Reber, 1993).

Further, contrary to our predictions, there was no significant difference in performance between the two complex sentence types, even though the Complex 2 pattern might be more cognitively taxing than the Complex 1 pattern. In fact, performance was slightly, though not significantly, better on Complex 2 sentences. Notably, responses on the debriefing questionnaire suggest that the VF-V1 pattern was particularly salient for learners, so benefits due to saliency (see van den Bos & Poletiek, 2008) may have overcome any limitations on learning due to additional cognitive complexity of the VF-V1 pattern.

Cognitive Abilities in Incidental and Instructed Learning

While we found correlations between cognitive abilities and learning outcomes, they accounted for little of the variance in overall performance. This is surprising, as previous research has shown a link between both WM (hypothesized to be more important in the explicit condition; e.g., Linck et al., 2014) and procedural learning (believed to play a more important role in implicit conditions; e.g., Morgan-Short et al., 2014) in L2 acquisition, and an interaction with implicit and explicit exposure conditions (Erlam, 2005; Robinson, 2002; Tagarelli et al., 2015). While there were some between-groups differences, they were in unexpected directions. The only significant correlation over all sentence types was between performance on the procedural learning task and the GJT in the incidental group, and this was a negative correlation. While some previous research suggests that cognitive IDs should be apparent under explicit, but not implicit, exposure conditions (de Graaff, 1997; Gebauer & Mackintosh, 2007; Reber et al., 1991), it is not particularly surprising that there were no correlations between learning outcomes and cognitive IDs in our explicit condition. Other studies have found no relationship between overall performance and IDs in explicit conditions (Grey et al., 2015), even when the ID in question was WM, which has been linked to explicit processes (e.g., Tagarelli et al., 2015). The provision of meta-linguistic information may actually be particularly effective in leveling out any influence of cognitive IDs (e.g., Erlam, 2005). The direction of the correlation between GJT performance and SRT scores in the incidental group is somewhat surprising, given that our measures of awareness demonstrate that these learners are relying on implicit knowledge in the GJT, and the SRT is a procedural/implicit learning task. We therefore expected these to be positively correlated, as procedural learning abilities have been shown to be positively related to L2 outcomes in implicit conditions, but only after more exposure and at higher proficiency (Morgan-Short et al., 2014).

It is worth noting that the measure of learning in this study is an untimed GJT. While oral untimed GJTs have been used to measure implicit knowledge (e.g., Bialystok, 1982), R. Ellis (2005) suggests that these tasks may indirectly tap explicit knowledge, as the untimed nature of the task allows learners to (i) attend to the meaning of the sentence, (ii) notice if there was a mistake and (iii) reflect on the mistake. Performance on this task might therefore be less related to procedural learning abilities than expected, even if participants were exposed in an implicit condition. Additionally, the use of a semi-artificial language in the present study may have shifted language processing toward more explicit mechanisms (cf. Godfroid, this issue). Participants in the present study were learning a foreign syntax with an L1 lexicon, which likely made processing for meaning somewhat automatic, thus freeing up resources to direct more attention to form.

Exposure conditions, Complexity, and Individual Differences

Our third research question was largely exploratory, as little is known about the three-way relationship between exposure conditions, complexity, and cognitive IDs. Our findings suggest that we still have a lot to learn. Exposure condition was a significant predictor in regression models for both Simple and Complex 1 sentences, and accounted for more of the variance for Simple sentences. This suggests that performance on both simple and complex structures is influenced by the type of exposure, but that this influence is greater for simple structures. These were, indeed, the only structures for which learners in the instructed condition significantly outperformed those in the incidental condition. Nevertheless, exposure condition explained very little of the variance (10-14%). When WM and procedural learning ability were added to the regression model, more variance was accounted for, but only marginally so, and exposure condition was still the only significant (or trending) predictor in these models. This suggests that IDs had very little, if any, influence on the learning of the Simple and Complex 1 patterns.

Interestingly, exposure condition was not a significant predictor in the regression model for the Complex 2 sentences, accounting for only about 4% of the variance. However, adding IDs to this model, particularly the procedural learning measure (SRT), accounted for significantly more variance, though it was still low overall (14%). This suggests that for this particular sentence pattern, the type of training that learners received was not important, but procedural learning abilities were. Correlational analyses showed that procedural learning abilities did not influence learning in the instructed group, but in the incidental group, learners with better procedural memory abilities performed worse. Though not a significant predictor in the regression model, RSpan scores were significantly positively correlated with *d'* scores on these structures in the incidental condition.

At first glance, these results are quite surprising, as both implicit learning and performance on complex sentences might be expected to be positively related to procedural learning abilities, and not necessarily related to WM abilities (but see Erçetin & Alptekin, 2013). We begin by discussing the findings related to WM. Results from the debriefing questionnaires suggest that the Complex 2 grammatical sentences, VF-V1, were very salient to learners, as they were able to provide more examples of these sentences than V2-VF sentences, and almost the same number as V2 sentences. Good WM might help participants learn more salient rules in a less structured condition, while less salient rules go by unnoticed and are therefore more difficult to acquire (N. Ellis, 2006). This ability to attend to salient rules may be a driving force in incidental learning (van den Bos & Poletiek, 2008), and may be more important when L2 learning is more naturalistic, as opposed to classroom-based, which could have implications for our incidental condition (DeKeyser, 2005). Indeed, knowledge of such salient rules is more likely to be explicit in nature, and so this finding is consistent with literature suggesting that high WM span is beneficial for explicit language processes.

Regarding the relationship between the SRT and learning outcomes for Complex 2 sentences in the incidental condition, we can refer back to our more general discussion of this relationship across all sentence types. The nature of the untimed GJT (R. Ellis, 2005), and perhaps of the semi-artificial language, as well as the saliency of the Complex 2 forms, may have favored more explicit processing of this pattern. Perhaps learners with good procedural memory abilities used this system to extract more complex structural information about these forms (see Ullman, 2015), while the learners with poorer procedural memory abilities instead relied on the more salient, explicit cues, which was ultimately more successful. It should be noted, though, that analyses of source attributions according to sentence type suggest that learners in the incidental condition were highly accurate when relying on implicit knowledge for the Complex 2 sentences (see Table S4.2). These results therefore may not be so clear, and may also call into question the reliability of self-reported awareness measures.

Finally, while we observed effects of all three variables in our analyses, regression models demonstrate that exposure condition and IDs only predict a very small amount of the variance in learning outcomes for all sentence types. This once more suggests that there is much more to the puzzle that was not investigated in this study.

Limitations

This study is not without limitations. First, there was a significant correlation between RSpan and SRTz scores in the incidental group, but not in the instructed group. The variance inflation factors (VIF) for the regression analyses are all close to 1, indicating that the estimation of the regression coefficients is not overestimated due to multicollinearity. Still, this finding reveals a difference between the groups, so results pertaining to IDs should be interpreted with caution. While we have a reasonable number of participants in each group, correlational analyses, particularly regressions, would benefit from larger sample sizes, in which between-groups differences such as these should not arise by chance.

Another potential limitation of this study is our operationalization of linguistic complexity. First, we defined structural complexity in terms of T-units, which is only one of many possible ways to do so, and perhaps not the most sensitive. Second, we used cognitive complexity, or difficulty, to distinguish the Complex 1 and Complex 2 sentences. Bulté and Housen (2012) highlight the subjectivity involved in assessing the difficulty of a particular language feature, as cognitive complexity is relative, or learner-dependent. Results from our debriefing questionnaires indeed suggested that Complex 2 sentences were quite salient to learners, and possibly not as difficult as we anticipated.

Some of our results call into question the reliability of using subjective self-reported awareness measures when trying to assess the nature of L2 knowledge (see also Rebuschat, 2013). For example, source attributions indicated that learners in the incidental condition were relying on implicit knowledge during the GJT, but the relationship between cognitive IDs and learning outcomes in this group suggest that explicit processes were at work.

Finally, using accuracy on the GJT as a measure of learning outcomes may have biased our results toward explicit processes. While it was beyond the scope of the current paper, we also collected data that may speak more to implicit processes during training and testing, including elicited imitations, accuracy and reaction time on the plausibility judgment task, and reaction time on the GJT. Further analyses on this data should help tease apart the role of implicit and explicit processes in L2 learning.

Conclusions

Overall, this study revealed an interesting relationship between exposure conditions, IDs, and syntactic complexity. Although learning effects were observed in both conditions, our findings clearly demonstrate the advantage of providing metalinguistic instruction prior to meaning-focused exposure. Subjective measures of awareness demonstrated that learners in the incidental group seemed to rely on implicit knowledge, whereas learners in the instructed group rely on explicit knowledge. Performance across three syntactic patterns of varying complexity showed that while explicit instruction led to overall better performance on all structures, both groups showed similar patterns regarding the structures on which they performed best (Simple, followed by Complex 2, then Complex 1). Finally, analyses of IDs demonstrated the importance of considering multiple measures of cognitive abilities when assessing the relationship between IDs and exposure conditions, as abilities may relate differently to learning within and across conditions. This study goes a step further by examining the extent to which this relationship is mediated by linguistic complexity. There are limitations that necessitate a cautious interpretation of these results, but overall this study serves as a timely reminder of the multiplicity of factors that affect the process of L2 learning.

[1] Participants also completed a nonword repetition task and a Stroop task, which are beyond the scope of this paper.

References

Alanen, R. (1995). Input enhancement and rule presentation in second language acquisition. In R. Schmidt (Eds.), *Attention and awareness in foreign language learning* (pp. 259-302). Honolulu: University of Hawai‘i, Second Language Teaching & Curriculum Center.

Baars, B. J. (1988). *A cognitive theory of consciousness*. New York: Cambridge University Press.

Baddeley, A. D. (2000). The episodic buffer: A new component of working memory? *Trends in Cognitive Sciences, 4*(11), 417-423.

Baddeley, A. D. (2010). Working memory. *Current Biology, 20*(4), R136-R140.

Baddeley, A. D., & Hitch, G. (1974). Working memory. In G.H. Bower (Ed.), *The psychology of learning and motivation* (pp. 47–90). New York: Academic Press.

Bialystok, E. (1982). On the relationship between knowing and using forms. *Applied Linguistics, 3*, 181-206.

Brandt, S., Diessel, H., & Tomasello, M. (2008). The acquisition of German relative clauses: A case study. *Child Language*, *35*, 325–348.

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

Bygate, M. (1999). Task as context for the framing, reframing and unframing of language. *System*, *27*, 33-48.

Clark, R. E., & Squire, L. R. (1998). Classical conditioning and brain systems: A key role for awareness. *Science*, *280*, 77–81.

Daneman, M., & Carpenter, P.A. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning and Verbal Behavior*, *19*, 450-466.

de Graaff, R. (1997). The eXperanto experiment: Effects of explicit instruction on second language acquisition. *Studies in Second Language Acquisition*, *19,* 249-276.

DeKeyser, R. M. (1995). Learning second language grammar rules: an experiment with a miniature linguistic system. *Studies in Second Language Acquisition*, *17,* 379–410.

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

Destrebecqz, A., & Cleeremans, A. (2001). Can sequence learning be implicit? New evidence with the process dissociation procedure. *Psychonomic Bulletin & Review*, *8,* 343–350.

Dienes, Z., & Scott, R. (2005). Measuring unconscious knowledge: distinguishing structural knowledge and judgment knowledge. *Psychological Research*, *69*, 338 - 351.

Dörnyei, Z. (2006). Individual differences in second language acquisition. *AILA Review*, *19*, 42-68.

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

Ellis, R. (2007). The differential effects of corrective feedback on two grammatical structures. In A. Mackey (Ed.), Conversational interaction in second language acquisition (pp. 339–60). Oxford, England: Oxford University Press.

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

Erlam, R. (2005). Language aptitude and its relationship to instructional effectiveness in second language acquisition. *Language Teaching Research*, *9*(2), 147-171.

Gaies, S. J. (1976). *Sentence-combining: A technique for assessing proficiency in a second language*. Paper presented at the Conference on Perspectives on Language, University of Louisville, Kentucky.

Godfroid, A. (in press). The effects of implicit instruction on implicit and explicit knowledge development. *Studies in Second Language Acquisition*, *38*(2).

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

Granena, G. (2015). Explicit and implicit cognitive aptitudes and information-processing styles: An individual differences study. *Applied Psycholinguistics*, *FirstView*, 1-14.

Grey, S., Williams, J. N., & Rebuschat, P. (2014). Incidental exposure and L3 learning of morphosyntax. *Studies in Second Language Acquisition*, *36*, 1-34.

Grey, S., Williams, J. N., & Rebuschat, P. (2015). Individual differences in incidental language learning: The role of learning styles, personality, and working memory. *Learning and Individual Differences*, *38*, 44-53.

Hamrick, P. (2014). A role for chunk formation in statistical learning of second language syntax: Chunk formation and L2 syntax. *Language Learning*, *64*(2), 247–278.

Hamrick, P. (in press). Declarative and procedural memory abilities as individual differences in incidental language learning. *Learning and Individual Differences.*

Han, Y., & Ellis, R. (1998). Implicit knowledge, explicit knowledge and general language proficiency. *Language Teaching Research*, *2*, 1-23.

Hawkins, J.A. (2004). *Efficiency and complexity in grammars*. Oxford: Oxford University Press.

Housen, A., Pierrard,M., & Van Daele, S. (2005). Rule complexity and the effectiveness of explicit grammar instruction. In A. Housen & M. Pierrard (Eds.), Investigations in Instructed Second Language Acquisition (pp. 235–70). Berlin, Germany: De Gruyter.

Howard, J.H. Jr., Howard, D.V. (1997). Age differences in implicit learning of higher-order dependencies in serial patterns. *Psychology and Aging*, *12*, 634–656.

Hummel, K. M. (2009). Aptitude, phonological memory, and second language proficiency in non-novice adult learners. *Applied Psycholinguistics*, *30*, 225–249.

Hunt, K. (1965). *Grammatical structures written at three grade levels*. NCTE Research report No. 3. Champaign, IL, USA: National Council of Teachers of English.

Jarvella, R.J., & Herman, S.J. (1972). Clause structure of sentences and speech processing. *Perception & Psychophysics*, *11*(5), 381-384.

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

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

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

Krashen, S. (1994). The input hypothesis and its rivals. In N. Ellis (Ed.) *Implicit and explicit learning of languages* (pp. 45-77). London: Academic Press.

Larsen-Freeman, D. (1978). An ESL index of development. *TESOL Quarterly*, *12*, 439-448.

Larsen-Freeman, D. (1983). Assessing global second language proficiency. In H. W. Seliger & M. H. Long (Eds.), *Classroom oriented research* (pp. 287-304). Rowley, MA: Newbury House.

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

Leung, J. H. C., & Williams, J. N. (2012). Constraints on implicit learning of grammatical form-meaning connections. *Language Learning*, *62*(2), 634–662.

Linck, J. A., Osthus, P., Koeth, J. T., & Bunting, M. F. (2014). Working memory and second language comprehension and production: A meta-analysis. *Psychonomic Bulletin & Review*, *21*, 861–883.

Lord, C. (2002). Are subordinate clauses more difficult? In J. Bybee & M. Noonan (Eds.), *Complex sentences in grammar and discourse*. Essays in honor of Sandra A. Thompson (pp. 223-233). Philadelphia: John Benjamins.

Martin, K. I. & Ellis, N.C. (2012). The roles of phonological short-term memory and working memory in L2 grammar and vocabulary learning. Studies in Second Language Acquisition, 34, 379-413.

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

Miyake, A., & Friedman, N.P. (1998). Individual differences in second language proficiency: Working memory as language aptitude. In Healy, A. & Bourne, L. (Eds.), *Foreign Language Learning* (pp. 339-364). Mahwah, NJ: Lawrence Erlbaum.

Morgan-Short, K., Sanz, C., Steinhauer, K., & Ullman, M. T. (2010). Second language acquisition of gender agreement in explicit and implicit training conditions: An event-related potential study. *Language Learning*, *60*(1), 154-193.

Morgan-Short, K., Finger, I., Grey, S., & Ullman, M. T. (2012a). Second language processing shows increased native-like neural responses after months of no exposure. *PLoS ONE*, *7*(3), e32974.

Morgan-Short, K., Steinhauer, K., Sanz, C., & Ullman, M. T. (2012b). Explicit and implicit second language training differentially affect the achievement of native-like brain activation patterns. *Journal of Cognitive Neuroscience*, *24*(4), 933-947.

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.

Morgan-Short, K., Den, Z., Brill-Schuetz, K., Faretta-Stutenberg, M., & Wong, F. (2015). A view of the neural representation of second language syntax through artificial language learning under implicit contexts of exposure. *Studies in Second Language Acquisition*, *37*, 383–419.

Nissen, M. J. & Bullemer, P. (1987). Attentional requirements of learning: Evidence from performance measures. *Cognitive Psychology*, *19*, 1–32.

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

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

Rebuschat, P. (2008). *Implicit learning of natural language syntax*. Ph. D. thesis, Cambridge University, Cambridge.

Rebuschat, P. (2013). Measuring implicit and explicit knowledge in second language research. *Language Learning*, *63*, 595-626.

Rebuschat, P., Hamrick, P., Riestenberg, K., Sachs, R., & Ziegler, N. (2015): Triangulating measures of awareness: A contribution to the debate on learning without awareness. *Studies in Second Language Acquisition*, *37*(2), 299 – 334.

Rebuschat, P., & Williams, J. N. (2012). Implicit and explicit knowledge in second language acquisition. *Applied Psycholinguistics*, *33*, 829–856.

Robinson, P. (1997). Individual differences and the fundamental similarity of implicit and explicit adult second language learning. *Language Learning*, *47*, 45–99.

Robinson, P. (2002). *Individual differences and instructed language learning*. Amsterdam: Benjamins.

Robinson, P. (2005). Aptitude and second language acquisition. *Annual Review of Applied Linguistics*, *25*, 46-73.

Roehr, K. (2008). Linguistic and metalinguistic categories in second language learning. *Cognitive Linguistics*, *19*, 67–106.

Rogers, J., Révész, A., & Rebuschat, P. (in press). Implicit and explicit knowledge of L2 inflectional morphology. *Applied Psycholinguistics*.

Scott, M. & Tucker, C.R. (1974). Error analysis and second language strategies. *Language Learning*, *24*, 69-97.

Shah, P., & Miyake, A. (1999). Models of working memory: An introduction. In A. Miyake & P. Shah (Eds.), *Models of working memory: Mechanisms of active maintenance and executive control* (pp. 1–27). Cambridge, UK: Cambridge University Press.

Shanks, D. R. (2005). Implicit learning. In K. Lamberts & R. Goldstone (Eds.), *Handbook of cognition* (pp. 202–220). London, UK: Sage.

Skehan, P. (1989). *Individual differences in second-language learning.* New York: Edward Arnold.

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

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

Squire, L. R., & Knowlton, B. J. (2000). The medial temporal lobe, the hippocampus, and the memory systems ofthe brain. In M. S. Gazzaniga (Ed.), The new cognitive neurosciences (pp. 765–780). Cambridge, MA: MITPress.

Tagarelli, K. M., Borges Mota, M. & Rebuschat, P. (2015). Working memory, learning context, and the acquisition of L2 syntax. In Zhisheng, W., Borges Mota, M. & McNeill, A. (Eds.), *Working memory in second language acquisition and processing: theory, research and commentary* (pp. 224–247). Bristol: Multilingual Matters.

Ullman, M.T. (2015). The declarative/procedural model: A neurobiologically motivated theory of first and second language. In B. VanPatten and J. Williams (Eds.), *Theories in second language acquisition: An introduction* (2nd ed.) (pp. 135-158). New York: Routledge.

Unsworth, N., Heitz, R. P., Schrock, J. C., & Engle, R. W. (2005). An automated version of the operation span task. *Behavior Research Methods*, *37*, 498-505.

Yang, L., & Givón, T. (1997). Benefits and drawbacks of controlled laboratory studies of second language acquisition. *Studies in Second Language Acquisition*, *19*, 173–194.

Williams, J. N. (2003). Inducing abstract linguistic representations: human and connectionist learning of noun classes. In R. van Hout, A. Hulk, F. Kuiken, & R. Towell (Eds.), *The interface between syntax and the lexicon in second language acquisition* (pp. 151–174). Amsterdam: Benjamins.

Williams, J. N. (2005). Learning without awareness. *Studies in Second Language Acquisition, 27*(2), 269-304.

Williams, J. N. (2012). Working memory and SLA. In S. M. Gass & A. Mackey (Eds.), *The Routledge handbook of second language acquisition* (pp. 427–441). New York, NY: Routledge.

Williams, J.N. & Lovatt, P. (2003). Phonological memory and rule learning. *Language Learning*, *53*, 67–121.

van den Bos, E. & Poletiek, F.H. (2008). Intentional artificial grammar learning: When does it work? *European Journal of Cognitive Psychology*, *20*(4), 793-806.

Table 1
Examples of grammatical and ungrammatical (*) sentence patterns. Sentences below are examples taken from the testing set.

Pattern	*Example sentence*
Simple	
V2	Yesterday *scribbled* David a long letter to his family.
*V3	Some time ago John *filled* the bucket with apples.
*V4	Recently Susan much furniture *imported* for her new weekend retreat.
Complex 1	
V2-VF	In the afternoon *acknowledged* David that her children to England *moved*.
*V1-VF	*Stayed* Jennifer at the hotel because her husband yesterday a boring conference *attended*.
*VF	After dinner Susan an old car with her savings *bought*.
Complex 2	
VF-V1	When his wife in the afternoon the office *left*, *prepared* Jim dinner for the entire family.
*VF- V2	Because his children recently a calculator *required*, Jim *called* the electronics store.
*V1	*Invited* Emma after dinner some colleagues to her birthday party.

Note: Ungrammatical sentences are matched to grammatical sentences based on features of verb placement, not complexity. This is why there are single-clause ungrammatical structures in the Complex 1 and Complex 2 groups.

Table 2

Accuracy (%) and *d'* scores on the GJT for each group overall and according complexity.

	Incidental				Instructed			
	Accuracy		*d'*		Accuracy		*d'*	
	M	SD	M	SD	M	SD	M	SD
All	55.53	9.69	0.19**	0.32	67.33	15.65	0.63****	0.64
Simple	60.60	13.49	0.42***	0.64	75.60	17.70	1.00****	0.81
Complex 1	50.00	11.82	0.04	0.49	63.12	18.77	0.46*	0.85
Complex 2	56.00	19.20	0.28^{+}	0.74	63.53	25.00	0.64***	0.92

Significance from chance: $^{+}p = .07$, *$p < .05$, **$p < .01$, ***$p < .005$, ****$p < .001$.

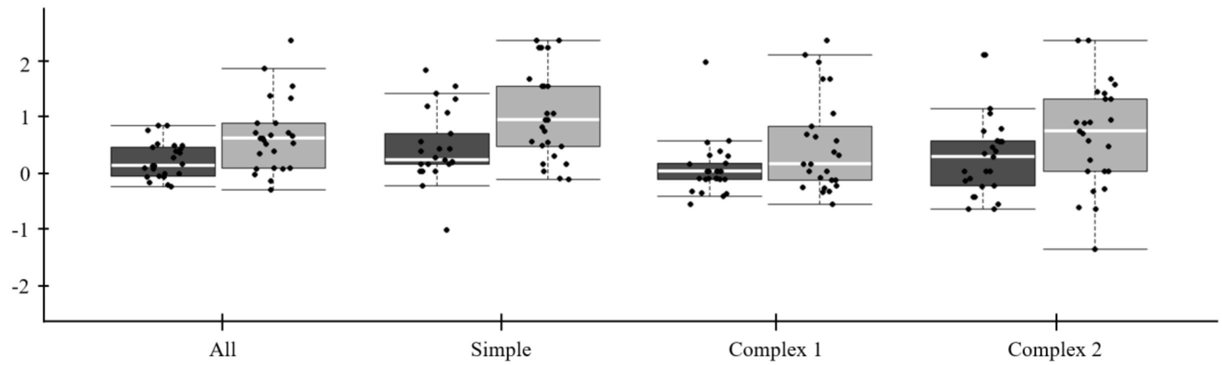

Figure 1. Boxplots of *d*-prime performance for all sentences, and for each sentence group. Error bars represent standard deviation. Dark grey boxes = incidental condition; light grey boxes = instructed condition.

Supplementary Materials

S1. Participant Information

Table S1.1
Participant information and scores on independent measures according to location (US v. UK) and exposure condition (Incidental v. Instructed).

		Incidental			***Instructed***		
	Score range or unit	*US*	*UK*	*Combined*	*US*	*UK*	*Combined*
Participants		12	13	25	11	14	25
Male		5	1	6	2	3	5
Female		7	12	19	9	11	20
Age Mean		19.00	19.54		18.55	19.93	
SD		1.05	1.13		1.04	1.54	
RSpan Mean	0-42	18.91	22.62		21.09	24.85	
SD		7.37	8.64		8.04	8.74	
ASRT Mean	0-18	12.50	--	--	12.80	--	--
SD		3.24	--	--	2.53	--	--
SRT Mean	ms	--	550	--	--	240	--
SD		--	430	--	--	251	--

Statistical analyses for each variable

Gender. A three-way loglinear analysis with location (UK vs. US), condition (incidental vs. instructed), and sex (male v. female) was conducted to determine whether these variables were evenly distributed. The likelihood ratio of this model was χ^2 (6) = 4.76, *p* = .575. The highest-order interaction (location X condition X sex) was not significant, χ^2 (1) = 2.50, *p* = .114, nor were any two-way interactions (all *p*s > .19). There was a significant effect of sex, χ^2 (1) = 16.62, *p* < .001, in which more females participated than males. In sum, participants were evenly distributed across incidental and instructed conditions in the US and the UK. Overall, there were more female than male participants, but there was no significant difference between the proportions of males and females in each bin (i.e., UK-incidental, UK-instructed, US-incidental, US-instructed).

Age. A 2x2 ANOVA with Condition x Location as factors revealed no main effect of condition, *F* < 1, and no group X condition interaction, *F*(1, 44) = 1.383, *p* = .246. There was a main effect of location, *F*(1, 44) = 7.160, *p* = .01. The participants in the UK were on average one year older than the participants in the US, and this difference was significant. However, at this age, the difference of one year should not make a difference in performance on the tasks in this study.

Readings Span. A 2x2 ANOVA with Condition x Location as factors revealed no significant main effects or interactions.

Main effect of Condition: *F* < 1

Main effect of Location: *F*(1, 44) = 2.413, *p* = .126

Condition X Location Interaction: $F < 1$

(A)SRT. An independent samples *t*-test revealed no differences between conditions, $t < 1$.

SRT. An independent samples *t*-test revealed significant differences between conditions, $t(24) = 2.248$, $p = .034$. However, tests of skewness and kurtosis for each group and variance between groups (Levene's Test of Equality of Variance) are all nonsignificant ($p > .05$), which demonstrate that the scores are normally distributed and have similar variance between groups, so correlational analyses can be conducted.

Based on the above analyses, we judged that it was appropriate to collapse participants from the US and UK, within their respective conditions. Additionally, there were no differences between conditions that should influence their performance on the GJT, independent of the variable that we manipulated, which was type of exposure.

Language Background

All participants were native speakers of English, but their additional language experience was quite varied. Two participants reported additional second languages, Russian and Mandarin. All but three participants had studied at least one second language (*range* = 0-4, *mean* = 1.41, *SD* = 0.85). Second languages spoken by the participants included French (*n* = 31), Spanish (*n* = 19), Mandarin (*n* = 9), Arabic (*n* = 3), Japanese (*n* = 2), Italian (*n* = 2), Welsh (*n* = 2), and Latin (*n* = 2).

S.2 Definitions for Source Attributions

Participants were asked to select the guess category when they believed the judgment to be based on a guess, i.e. they might as well have flipped a coin. If they were somewhat confident in their decision but did not know why it was right, they were supposed to select the intuition category. The memory category was designated for judgments based on the conscious recollection of entire sentences (or parts of sentences) from the training phase. Finally, participants were asked to select the rule category if the judgment was based on a rule that was acquired during the training phase and that they would be able to verbalize at the end of the experiment. All participants were provided with these definitions before starting the testing phase.

S3. Subjective measures of awareness

Table S3.1

Accuracy and proportions (%) across source attributions for all sentence types

		Guess	*Intuition*	*Memory*	*Rule*
Incidental					
Accuracy	Mean	49.1	56.8*	50.9	50.7
	SD	29.2	13.9	34.7	21.1
Proportion		12.5	41.5	15.5	30.5
Instructed					
Accuracy	Mean	50.0	53.6	55.9	68.8***
	SD	32.3	19.9	30.8	26.9
Proportion		6.1	26.3	24.2	43.5

Significance from chance: * $p < .05$, *** $p < .005$, **** $p < .001$

Table S3.2
Accuracy and proportions (%) across source attributions by complexity

		Guess	Intuition	Memory	Rule
Incidental					
Simple 1					
Accuracy	Mean	19.0	80.6	53.9	47.2
	SD	18.8	17.3	50.4	41.1
Proportion		10.8	39.2	13.8	36.2
Complex 1					
Accuracy	Mean	23.0	50.0	63.0	52.6
	SD	22.3	16.7	37.5	19.4
Proportion		10.6	38.2	19.0	32.2
Complex 2					
Accuracy	Mean	42.0	90.5*	67.0	42.9
	SD	36.7	16.5	57.7	51.5
Proportion		16.0	47.2	13.8	23.0
Instructed					
Simple 1					
Accuracy	Mean	50.0	66.7*	87.5	100****
	SD	7.1	0.0	17.7	0
Proportion		7.0	24.4	23.4	45.2
Complex 1					
Accuracy	Mean	25.0	26.8	55.0	54.2*
	SD	35.4	2.5	7.1	5.9
Proportion		5.6	26.5	26.0	41.9
Complex 2					
Accuracy	Mean	17.0	60.0	90.0***	90.0^{+}
	SD	23.6	56.6	14.1	14.1
Proportion		5.6	28.0	23.0	43.4

Significance from chance: $^{+}$ $p < .1$, * $p < .05$, *** $p < .005$, **** $p < .001$

Mixed-effects repeated measures ANOVA to determine the effect of condition, sentence group, and source attribution on accuracy:
Main effects:
No main effect of Sentence Group: $F(2, 6) = 2.064$, $p = .208$, partial $\eta^2 = .408$, observed power = .276.
Main effect of SA: $F(3, 9) = 7.674$, $p = .008$, partial $\eta^2 = .719$, observed power = .910.

No main effect of Condition: $F < 1$

Interactions:
No Sentence Group X Condition interaction: $F(2, 6) = 1.182$, $p = .369$, partial $\eta^2 = .283$, observed power = .175.
No Sentence Group X SA interaction: $F < 1$
Trending SA X Condition interaction: $F(3, 9) = 3.273$, $p = .073$, partial $\eta^2 = .522$, observed power = .548. (large effect size/power)
No 3-way interaction: $F < 1$

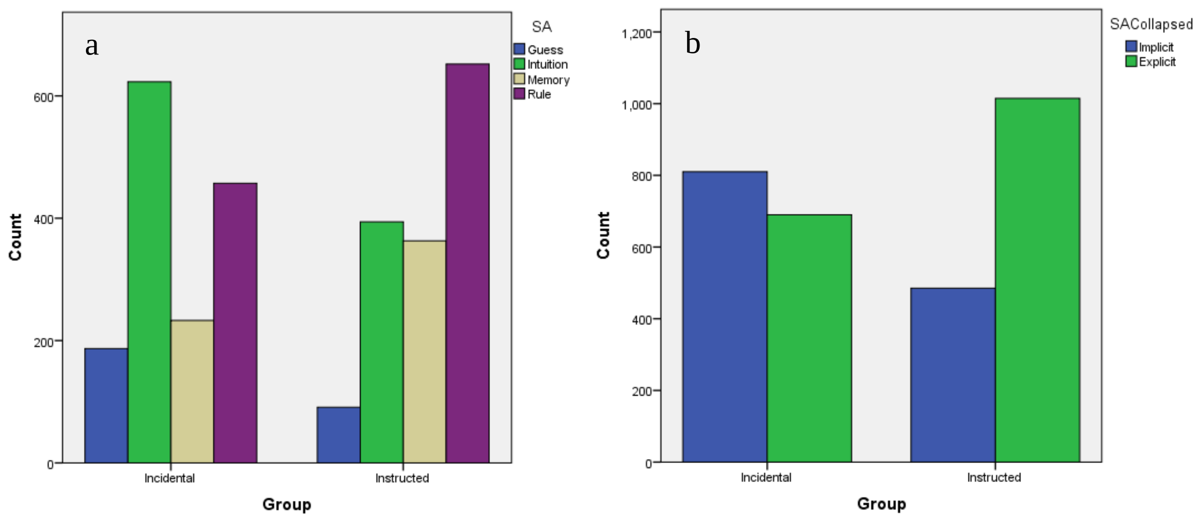

Figure S3.1. (a) Distribution of all source attributions across the two exposure conditions and (b) distribution of source attributions collapsed into implicit (guess and intuition) and explicit (memory and rule) categories.

Table S3.3
Accuracy and proportions (%) across confidence ratings over all sentence groups

		Not at all confident	Somewhat confident	Quite Confident	Extremely Confident
Incidental					
Accuracy	Mean	54.2	54.1^{+}	57.8*	52.3
	SD	31.8	11.2	15.3	24.7
Proportion		6.5	40.9	37.9	14.7
Instructed					
Accuracy	Mean	49.4	54.8	66.3****	70.7****
	SD	37.5	21.9	21.2	25.6
Proportion		4.5	26.7	39.0	29.9

Significance from chance: $^{+}$ $p < .1$, * $p < .05$, *** $p < .005$, **** $p < .001$

Mixed-effects repeated measures ANOVA to determine the effect of condition and confidence rating on overall accuracy:

No main effect of CJ on Accuracy: $F(3, 84) = 1.152$, $p = .333$
No main effect of Group on Accuracy: $F(1, 28) = 1.626$, $p = .213$
No CJ X Group interaction: $F(3, 84) = 1.578$, $p = .201$

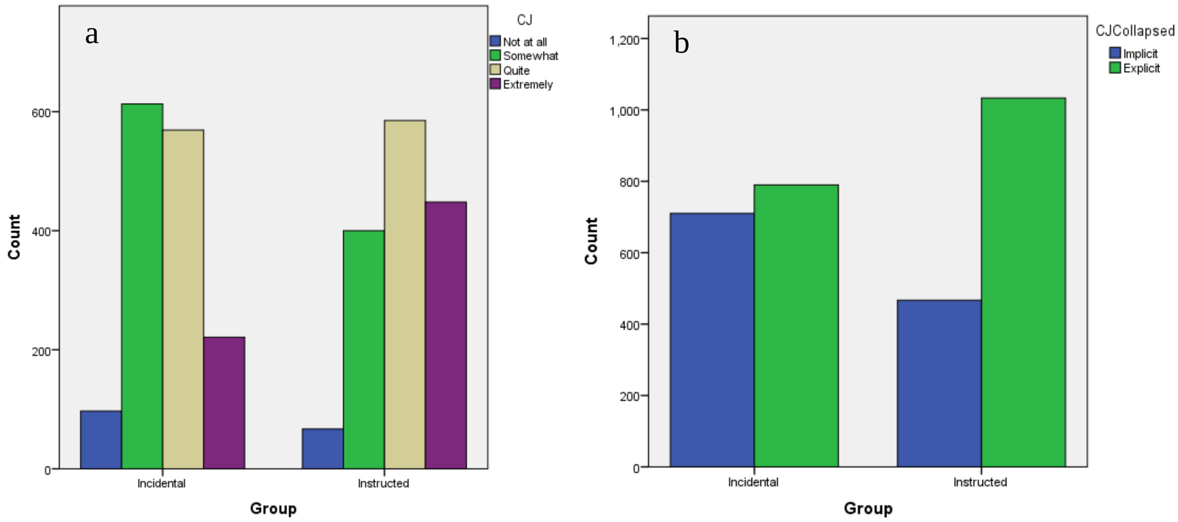

Figure S3.2. (a) Distribution of all confidence ratings across the two exposure conditions and (b) distribution of confidence ratings collapsed into implicit (guess and intuition) and explicit (memory and rule) categories.

Table S3.4
Accuracy and proportions (%) across confidence ratings by complexity

		Not at all confident	Somewhat confident	Quite Confident	Extremely Confident
Incidental					
Simple 1					
Accuracy	Mean	42.0	43.2	69.7****	70.4
	SD	41.9	17.0	14.0	27.1
Proportion		6.8	35.4	37.6	20.2
Complex 1					
Accuracy	Mean	50.0	73.7	55.8	75.0
	SD	40.1	23.9	23.3	50.0
Proportion		4.8	40.0	41.0	14.2
Complex 2					
Accuracy	Mean	13.0	63.8^{+}	61.7^{+}	62.5
	SD	25.0	29.2	37.6	47.9
Proportion		7.8	47.2	35.2	9.8
Instructed					
Simple 1					
Accuracy	Mean	33.0	68.5*	87.5****	97.0****
	SD	28.9	27.4	21.7	5.2
Proportion		4.4	25.2	37.8	32.6
Complex 1					
Accuracy	Mean	19.0	62.5	50.0^{+}	72.2*
	SD	33.0	23.2	22.0	19.2
Proportion		6.2	26.5	43.0	24.3
Complex 2					
Accuracy	Mean	11.0	49.8	66.7*	48.9^{+}
	SD	19.2	14.6	33.3	42.9
Proportion		2.6	28.4	35.7	33.3

Significance from chance: $^{+}$ $p < .1$, * $p < .05$, *** $p < .005$, **** $p < .001$

Mixed-effects repeated measures ANOVA to determine the effect of condition, sentence group, and confidence rating on accuracy:

Main effects:

No main effect of Sentence Group: $F(2, 10) = 1.843$, $p = .208$, partial $\eta^2 = .269$, observed power = .296.

Main effect of CJ: $F(3, 15) = 9.415$, $p = .001$, partial $\eta^2 = .653$, observed power = .983.

No main effect of Condition: *F*

Interactions:
No Sentence Group X Condition interaction: $F(2, 10) = 1.395$, $p = .292$, partial $\eta^2 = .218$, observed power = .233.
No CJ X Condition interaction: $F < 1$
No Sentence Group X CJ interaction: $F < 1$
No 3-way interaction: $F < 1$

S4. Correlations between GTJ and ID measures

Table S4.1
Average scores and standard deviations for each condition on the RSpan and SRT, and significance of independent samples *t*-tests comparing the conditions.

		RSpan	SRT	ASRT	SRTz
Incidental	*Mean*	20.92	550	12.50	0.21
	SD	8.13	430	3.24	1.13
	Range	7-35	-260-1440	7-18	-1.99-2.76
Instructed	*Mean*	23.13	239	12.80	-.20
	SD	8.47	251	2.52	.77
	Range	10-39	-220-640	8-16	-1.64-1.18
p		.361	.034	.820	.156

Note. Scores for each task were computed as follows: RSpan = total number of items in correctly recalled sets, maximum of 42; SRT = reaction time in milliseconds in the random block minus in the last sequence block; ASRT = total number of blocks in which reaction time for sequence trials was less than reaction time for random trials, maximum of 18; SRTz = z-scores of SRT and ASRT

Table S4.2
Correlations between *d'* scores on the GJT (overall, and for Simple, Complex 1, and Complex 2 sentence groups) and individual difference measures (RSpan, SRTz).

	Simple	*Complex 1*	*Complex 2*	*RSpan*	*SRTz*
Both Groups					
All	.752****	.744****	.710****	.192	-.348*
Simple		.488****	.277^{+}	.085	-.242
Complex 1			.256^{+}	.224	-.173
Complex 2				.190	-.337*
RSpan					-.307*
Incidental					
All	.635****	.411*	.723****	.176	-.586**
Simple		.037	.179	-.061	-.244
Complex 1			.019	-.062	-.205
Complex 2				.431*	-.543**
RSpan					-.501*
Instructed					
All	.757****	.804****	.710****	.152	-.154
Simple		.606****	.242	.113	-.120
Complex 1			.298	.346^{+}	-.061
Complex 2				-.035	-.076
RSpan					-.051

Note: $^{+}$ $p < .1$, * $p < .05$, ** $p < .01$, *** $p < .005$, **** $p < .001$, uncorrected Bonferroni corrected $\alpha = .003$ for each group.

S5. Regression analysis

Table S5.1

Regression models examining learning condition, working memory, procedural learning, and performance on the GJT for Simple, Complex 1, and Complex 2 sentences.

	Simple			Complex 1			Complex 2		
Variable	*B*	*SEB*	*β*	*B*	*SEB*	*β*	*B*	*SEB*	*β*
Step 1									
Constant	.345	.162		.017	.154		.321	.185	
Learning Condition	.607	.226	.378*	.468	.216	.314*	.352	.259	.203
R^2		.143			.099			.041	
$F(43)$		7.184*			4.718*			1.852^{ns}	
Step 2									
Constant	.516	.361		-.349	.340		.274	.396	
Learning Condition	.564	.232	.351*	.426	.219	$.286^{+}$	.249	.255	.144
Procedural Learning	-.141	.128	-.167	-.039	.121	-.049	-.285	.141	-.311*
Working Memory	-.003	.015	-.030	.017	.014	.190	.004	.016	.037
R^2		.168			.142			.143	
$F(42)$		2.763^{+}			2.267^{+}			3.464*	

Note: $^{+}$ $p < .1$, * $p < .05$