The Effect of Training Condition and Working Memory on Second Language Development of a Complex Form: The Spanish Subjunctive

Briana Villegas and Kara Morgan-Short

1. Introduction

An important issue in second language (L2) acquisition is understanding the best conditions for the development of particular L2 forms. Additionally, it is important to know whether the benefits of certain conditions vary among learners based on individual differences. A large body of research has examined the benefits of learning under explicit or implicit conditions, and different meta-analyses of this research suggest that although implicit conditions lead to development, explicit conditions generally lead to more development (Goo, Granena, Yimaz & Novella, 2015; Norris & Ortega, 2000; Spada & Tomita, 2010). However, methodological limitations in the body of empirical research have not yet been fully addressed (Doughty, 2003; Sanz & Morgan-Short, 2005). In addition, there has been a call for investigations into the interaction between training conditions and individual differences (DeKeyser, 2012; Sanz, 2005), as well as more research with complex L2 forms (Housen & Simoens, 2016). The current study addresses these issues by extending Fernández (2008) who found benefits for the provision of grammatical information in a processing instruction (explicit) condition as compared to a structured input (implicit) condition for L2 development of the Spanish subjunctive (a complex form). However, it was beyond the purposes of that study to examine performance on posttests or to examine individual differences. Below we provide a brief review of the literature in order to motivate our study that examines the effects of explicit and implicit conditions on the development of a complex L2 form and how development may vary based on individual differences in working memory.

2. Review of literature

2.1. Training condition

Regarding the question of whether it is more effective to learn under explicit conditions, the consensus from three meta-analyses is that explicit conditions, in which rule presentation or direction to attend to form is provided, generally convey advantages over implicit conditions, which expose learners to target forms but do not present rules or direction to attend to form (Goo et al., 2015; Norris & Ortega, 2000; Spada & Tomita, 2010), although implicit conditions also lead to development. However, the body of empirical research that has been included in these meta-analyses has undergone a number of methodological critiques, which have brought into question the validity and finality of their conclusions (e.g., Doughty, 2003; Sanz & Morgan-Short, 2005). These critiques include unbalanced training conditions that favor explicit conditions both in regard to the amount and time of exposure as well as whether the type of practice is matched between conditions. When these issues have been addressed, advantages for explicit training are not always evidenced for behavioral or brain-based processing measures (e.g., Morgan-Short, Steinhauer, Sanz, & Ullman, 2012; Morgan-Short, Sanz, Steinhauer, & Ullman, 2010; Sanz & Morgan-Short, 2004). Thus, more research with balanced designs is called for and should be extended to address different types of linguistic structures (i.e., complex forms) and/or individual differences (DeKeyser, 2012).

* Briana Villegas, University of Illinois at Chicago, bville4@uic.edu. Kara Morgan-Short, University of Illinois at Chicago. An earlier version of this paper was presented at the 36th Annual Second Language Research Forum (SLRF), Columbus, OH, 2017. We are very grateful to the members of the Cognition of Second Language Acquisition (CogSLA) Laboratory for their support and assistance with data collection and to the reviewers who provided valuable feedback. Correspondence concerning this article should be addressed to Briana Villegas

2.2. Effects of training condition on complex forms

Interest in the effect of training conditions on complex forms, regardless of the definition of complexity that is adopted (e.g., linguistic structure, learnability, etc.: see Section 4.2), is motivated by two competing hypotheses (Hulstijn & de Graaff, 1994; Reber, 1993; Tagarelli, Ruiz, Vega, & Rebuschat, 2016). It has been predicted that complex rules may be too complicated for the learner to derive from the input without the help of explicit instruction (Hulstijn & de Graaff, 1994). The opposite has also been predicted: Reber (1993) posited that implicit instruction may be more beneficial for complex rules/forms that cannot be clearly or easily explained to learners. However, results from laboratory and classroom research have been mixed. Several studies (e.g., Housen, Pierrard, & Van Daele, 2005; J. Williams & Evans, 1998) provide evidence that complex forms (e.g., French and English passive voice) can be acquired from explicit instruction, but they do not reveal an advantage for explicit over implicit instruction. Other studies (e.g., DeKeyser, 1995; Robinson, 1996) hint at qualitative advantages for implicit learning, such as, participants being able to verbalize a complex rule after implicit training versus a simple rule after explicit training (Robinson, 1996) or descriptive advantages for explicit conditions (DeKeyser, 1995). In order to reach a more robust conclusion about the effect of explicit and implicit conditions on complex L2 forms (as well as on simple L2 forms) Spada and Tomita (2010) conducted a meta-analysis on the interaction between instruction type and type of language feature. They concluded that explicit instruction was more effective than implicit instruction for both complex and simple rules. However, the studies reviewed by Spada and Tomita (2010) generally have methodological issues related to the balanced nature of their explicit and implicit conditions, as noted above. For example, in Robinson (1996), the participants in the instructed (i.e., most explicit) condition received additional exposure in the form of a written version of the rules that could be referred to throughout the training session. In Dekeyser (1995), the implicit condition received meaningful practice while the explicit condition received grammar instruction in addition to meaningful practice.

One study that addressed a limitation often found in previous research on complex L2 forms, specifically whether practice is meaningful and matched between explicit and implicit conditions, is Fernández (2008). Fernández (2008: Experiment 2) looked at whether explicit information contributed to greater accuracy during practice with a complex form (Spanish subjunctive) than was found for practice alone. She tested this within the Processing Instruction (PI) paradigm (e.g., VanPatten, 1996; VanPatten & Cadierno, 1993). As such, the more explicit group included all the elements of PI, i.e., explicit information (EI) about the form and processing strategies followed by structured input (SI) practice, which is a task containing input that is structured to require learners to process the target form to be able to accurately complete the task. The more implicit group completed the SI practice with no EI or pre-practice exposure to the form. Results indicated that during the SI practice task, the explicit group started processing the complex form correctly faster than the implicit group and that they had higher accuracy scores overall. Thus, Fernández (2008) concluded that explicit information was beneficial for learning a complex form. However, this study may actually be detecting a benefit for pre-practice exposure to the target form regardless of the type of exposure (e.g., explicit or implicit). Thus, to better understand the role of explicit conditions on the acquisition of complex forms, it is necessary to provide each condition with balanced exposure to the form in terms of the number of exemplars and the length of pre-practice exposure. In addition, it is important to extend this question beyond practice to assessments of L2 development, as there is some evidence that initial advantages for explicit information dissipate over time (VanPatten, Collopy, Price, Borst, & Qualin, 2013).

2.3. Training conditions and individual differences

The effect of different training conditions on the development of L2 complex forms may be mediated by individual differences. One individual difference that has been shown to be related to L2 development is working memory (WM) (Linck, Osthus, Koeth & Bunting, 2014), which is, “the cognitive system(s) responsible for the control, regulation, and active maintenance of information in the face of distracting information” (p. 862). Among the few studies that have examined interactions between training conditions and WM, Sanz, Lin, Lado, Stafford, and Bowden (2014) used the PI/SI framework to examine effects of explicit and implicit conditions. They found that WM positively predicted results for the more implicit SI condition that did not receive metalinguistic information, but...
not for the more explicit PI condition that did. This suggests that any benefit of WM may be neutralized when learners are provided with explicit information. Similar results were found in Robinson (2002) who found WM to be positively correlated with incidental (i.e., more implicit) learning outcomes on a listening grammaticality judgment task. In contrast, other research has found a positive effect of WM for accuracy on grammatical items in an explicit rule-search condition and no effect in implicit or incidental conditions (Tagarelli, Borges-Mota, & Rebuschat, 2011). Still others have found no effect for working memory on acceptability judgment tasks in incidental or instructed conditions (e.g., Grey, Williams, & Rebuschat, 2015) even when complexity was taken into account (Tagarelli et al., 2016). In summary, WM effects have not always been present; but when present, they have varied by condition type and linguistic form. Thus, more research is necessary to tease apart the role WM plays in the development of complex forms under different training conditions.

3. The study

The goal of the present study is to extend Fernández (2008: Experiment 2) by examining the effects of explicit and implicit training conditions on the development of a complex form (i.e., Spanish subjunctive) within a PI/SI paradigm. Our study builds on previous research by controlling for methodological biases. First, the design reported here controls for exposure, such that, the explicit and implicit training conditions were balanced for both the amount of time and the number of examples shown before practice. Second, the explicit and implicit conditions were provided with the same meaningful practice. Thus, the conditions only differed in the type of exposure they received (explicit or implicit). We also included a control condition that received the same amount of exposure as the experimental groups in order to examine whether any effects of the training conditions were above and beyond effects of exposure to the form. Finally, post- and delayed post-tests were included to examine L2 development over time. In addition to methodological improvements, a measure of WM as an individual difference was included in order to address the following research questions:

RQ 1: Do explicit and implicit training conditions differentially affect the development of knowledge of an L2 complex form at immediate and delayed testing?

RQ 2: Do individual differences (i.e., WM) relate differently to L2 knowledge development of a complex L2 form under different training conditions?

4. Methods

4.1. Participants

Eighty-five L2 learners of Spanish were recruited from second-semester, university-level basic Spanish language courses, and were pseudo-randomly assigned to the explicit condition (EXP) \((N = 28)\), to the implicit condition (IMP) \((N = 29)\), or to the control plus exposure condition (C+) \((N = 28)\). In order to control for the effects of previous exposure to the target form, participants were recruited before they had been exposed to the target form in their class, and we confirmed their lack of prior knowledge of the form through a debriefing questionnaire. In addition, participants were excluded from the final analysis if they had more than 3 years of Spanish instruction before attending university (EXP: \(N = 1\); IMP: \(N = 1\); C+: \(N = 3\)), if they demonstrated above 60 percent accuracy (following Fernández, 2008) on target items on the pretest (EXP: \(N = 3\); IMP: \(N = 2\); C+: \(N = 1\)), or if they did not complete both testing sessions (EXP: \(N = 4\); IMP: \(N = 2\); C+: \(N = 1\)). For the EXP condition, the final group consisted of 19 participants\(^1\) (11 female), the IMP condition consisted of 24 participants (11 female), and the final C+ condition consisted of 23 participants (18 female). A series of one-way between subjects analysis of variance tests indicated that the participants in these conditions did not differ based on age, \(F(2,63) = .67, p = .51\), age of first exposure to Spanish, \(F(2,56) = .74, p = .48\), the number of languages spoken other than English and Spanish, \(F(2,36) = 0.14, p = .87\), or on their WM composite z-scores, \(F(2,59) = 0.09, p = .92\) (see Table 1). However, scores on a Spanish vocabulary test that assessed participants’ knowledge of the verbs used in the experiment did reveal differences, \(F(2,63) = 3.21, p = .05\). In

\(^1\) One more EXP participant was excluded due to a researcher error in test administration.
particular, the C+ condition scored significantly lower than the IMP condition. Similarly, there was a difference on pre-test scores for subjunctive items, $F(2,63) = 4.46, p = .02$, such that the C+ condition scored significantly lower than the EXP condition.

Table 1

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>C+ ($N = 23$)</th>
<th>EXP ($N = 19$)</th>
<th>IMP ($N = 24$)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>M (SD)</strong></td>
<td><strong>M (SD)</strong></td>
<td><strong>M (SD)</strong></td>
<td></td>
</tr>
<tr>
<td>Age</td>
<td>19.70 (2.82)</td>
<td>20.74 (3.03)</td>
<td>20.33 (3.00)</td>
</tr>
<tr>
<td>Age of exposure to Spanish</td>
<td>15.16 (4.34)</td>
<td>14.41 (5.46)</td>
<td>13.43 (4.08)</td>
</tr>
<tr>
<td>Additional languages spoken</td>
<td>1.18 (0.40)</td>
<td>1.20 (0.41)</td>
<td>1.27 (0.47)</td>
</tr>
<tr>
<td>WM composite Z-Score</td>
<td>0.38 (0.67)</td>
<td>-0.04 (0.85)</td>
<td>0.05 (0.84)</td>
</tr>
<tr>
<td>Vocabulary test</td>
<td>87.07 (8.97)</td>
<td>90.01 (9.16)</td>
<td>93.59 (8.45)</td>
</tr>
<tr>
<td>SBJV pre-test accuracy</td>
<td>32.85 (16.23)</td>
<td>46.84 (10.57)</td>
<td>41.85 (17.89)</td>
</tr>
</tbody>
</table>

**Note:** C+ = control plus condition, EXP = explicit condition, and IMP = implicit condition. SBJV = subjunctive items. WM = working memory.

4.2. Target structure

The present study is interested in factors that influence learning of complex L2 structures. Following Fernández (2008) and other previous PI research (Farley, 2001a; 2001b; 2004), we examine the third person singular Spanish subjunctive in expressions of doubt—e.g., *No creo que-[DOUBT] baile-
[3sgSBJV]todos los días* (‘I do not believe that she/he/you dance(s) everyday’). This form can be considered complex by many proposed definitions of complexity, including being learned late in the L2 (e.g., Pienemann, 1998); having marked linguistic features such as low saliency and low communicative value (see DeKeyser, 2005; Doughty & Williams, 1998; Spada & Lightbown, 2008); being identified by teachers as difficult for learners (e.g., Robinson, 1996), and requiring more derivational steps in the L2 than L1 to arrive at the correct form (Hulstijn & de Graaff, 1994). Indeed, the Spanish subjunctive a) is learned late (Contreras Aedo & Ferreira Cabrera, 2013), b) is considered to be difficult for L1 English-L2 Spanish learners (e.g., Collentine, 1995), c) lacks saliency and communicative value (VanPatten, 2002; 2004), and d) the indicative/subjunctive alternation is not present in expressions of doubt in the L1 (English) of the participants (Kempchinsky, 2009).

4.3. Materials and procedures

4.3.1. Training conditions

Two experimental training conditions, EXP and IMP, and a control, C+, condition were employed in the current study. Each condition consisted of exposure to the Spanish subjunctive followed by practice. Concerning exposure, the EXP condition received explicit information (EI) about the Spanish subjunctive. The EI provided in the present study was the same as the EI provided in Fernández (2008) and originally developed by Farley (2000). It consisted of information about how the form is conjugated, that it is triggered by expressions of doubt, and that it is often overlooked by learners due to its semantic redundancy. Within the EI, participants were exposed to 13 subjunctive verbs either in isolation or in example sentences. Participants were asked to read the EI, and the exposure lasted approximately 3 minutes. The IMP and C+ conditions received exposure that consisted of 13 meaningful question and response dialogues, that participants were asked to read, containing the same number of subjunctive verb examples as the EXP condition (i.e., 13) and the same length of exposure to these forms (i.e., ~3 minutes), see (1). However, no EI was provided.

Jill sing well? No think that sing well.  
‘Does Jill sing well? No, I do not think she sings well.’
After exposure, participants in experimental conditions completed practice with the target form. The practice was adopted directly from Fernández (2008) and consisted of SI. Specifically, the SI interpretation task consisted of 30 target items (21 subjunctive, 9 indicative) probing the relationship between verb mood with statements of doubt and statements of certainty, see (2) and (3). Participants were instructed to listen to the second half of a Spanish sentence that was presented aurally and then to choose between two possible sentence beginnings that were visually displayed on a computer screen, see (4). The only modification made to these materials was that the sentences were re-recorded so that the native, Spanish speaker voice that was heard in practice would be the same as the voice in the assessment tasks that we added to the experimental design (see Section 4.3.2).

(2) No pienso que-[DOUBT] juegue-[3sgSBJV] tenis todos los días.
   'I don’t think he/she/you play(s) tennis every day’

(3) Creo que-[CERTAIN] tiene-[3sgIND] una esposa bonita.
   'I think he/she/you has(have) a beautiful wife’

(4) Aurally presented sentence ending: [“…coma-[3sgSBJV] pollo a veces “]
   Sentence beginning options presented on the screen:
   A) No es verdad que-[DOUBT]…  B) *Es verdad que-[CERTAIN]…

Participants in the C+ condition, which was not a condition in Fernández (2008), did not receive practice but were exposed to the same 30 items that appeared in the SI practice. In this case, participants were instructed to listen to the full sentence in Spanish and to select one of two pictures that best reflected the meaning of the sentence. The images were carefully selected so as not to direct participants’ attention to the subjunctive/indicative distinction (see Figure 1). It is also important to note that all three training conditions included feedback after each practice item. The feedback consisted of: “Correct!” or “Incorrect!” along with their percent accuracy on the task. Accuracy on the interpretation task used for practice was determined by awarding participants one point for each correct response to target items for a total possible score of 30 points (21 subjunctive, 9 indicative) and is reported as percent accuracy.

Figure 1. Example images from the control plus training. Participants listened to the Spanish sentence and then had to select the image that was most related to the sentence they heard.

4.3.2. Tests of L2 knowledge

To measure L2 knowledge of the Spanish subjunctive, the current study adopted the interpretation task used in Farley (2000) as pre-, post-, and delayed post-tests. A total of 36 target items had been developed and were distributed among three versions of the interpretation task such that each version consisted of 24 items (12 target; 12 distractors). The target items included nine expressions of doubt and
three expressions of certainty, all probing the subjunctive/indicative distinction, see (5) and (6). The 12 distractor items were completely unrelated to the target form, see (7).

(5) No es verdad que-[DOUBT] coma-[3sgSBJV] pollo a veces
No is true that eat(s) chicken sometimes
‘It is not true that he/she/you eat(s) chicken sometimes’

(6) Es evidente que-[CERTAIN] canta-[psgIND] en el coche
Is evident that sing(s) in the car
‘It is evident that he/she/you sing(s) in the car’

(7) El libro-[sg] (*Los libros-[pl]) sobre los gatos es-[sg] muy interesante
The book (*the books) about the cats is very interesting
‘The book about the cats is very interesting’

The three versions were assigned to participants using a Latin square design such that each participant received a different version for their pre-, post-, and delayed post-tests. Like in the SI practice, during the pre-, post-and delayed post-test 2 participants were instructed to listen to the second half of a Spanish sentence and then to choose between two possible visually displayed beginnings. No feedback was provided during any of these tasks. Accuracy on the interpretation task used for pre-, post-, and delayed post-tests was determined by awarding participants one point for each correct response to target items for a total possible score of 12 points (9 subjunctive, 3 indicative) and is reported as percent accuracy.

4.4. Working memory

Working memory was assessed with complex span tasks that required participants to both store and process information. To maximize validity while also managing time constraints, shortened operation, reading and symmetry span tasks were utilized to assess participants’ WM capacity (Oswald, McAbee, Redick, & Hambrick, 2015). In the operation span task, participants were instructed to remember or store a series of letters that were presented in sequences ranging from 3 to 7. In between the presentation of each letter, participants had to complete a processing task. In this case, they had to solve a simple math problem. After participants had seen all the letters and math problems in a given sequence, they were instructed to report which letters they saw and the order in which they saw them. The reading span task followed the same structure as above, but participants were presented with a different processing task. Instead of solving math problems between letters, the learners had to make semantic plausibility judgments about English sentences (e.g., Not plausible: “The man likes to run in the spaghetti”; Plausible: “The man likes to run in the park”). The symmetry span task again followed the same structure as the previous two tasks. However, the storage items and processing task were different. The storage items in this task were the sequence and location of red squares that appeared in a 4x4 grid, and the processing task required participants to determine whether images were symmetrical across their horizontal axis or not. As in the previous tasks, at the end of each sequence participants were instructed to report the location of each square that they had seen (i.e., the storage items) in the order in which they saw them. For all the working memory span tasks, participants’ accuracy and errors were recorded for both storage and processing items. A partial-scoring method was used, such that, participants received one point for each storage item recalled in the correct order (Conway et al., 2005). Following Conway et al. (2005), we only included scores that were at or above 80% accuracy on the processing items to ensure that participants were engaging in both the storage and processing components. Participants who scored below 80% on two or more of the three tasks were excluded from the analysis entirely (EXP: N

2 In the post and delayed post-test participants were prompted to provide confidence ratings and source attributions. Confidence ratings and source attributions were not included in the pre-test to avoid participants becoming aware of the target structure. Additionally, confidence ratings and source attributions will not be reported in the current paper.
Finally, the included scores from each complex span task were converted to z-scores and averaged to form a composite WM z-score.

4.5. Procedure

This study was carried out over two sessions separated by two weeks ($M = 15.72$ days, $SD = 4.03$). During the first session, participants completed a consent form, background questionnaire, and a vocabulary quiz. The vocabulary quiz was based on a handout of Spanish verbs in the infinitive forms that was provided to students by email before their testing session. This quiz was included so as to control for any differences in performance due to variable knowledge of the target verb meanings. After the vocabulary quiz, the pre-test was administered, followed by the training condition, and then the immediate post-test. Once these three tasks were completed, the participants were scheduled for their follow-up session. In the second session, participants signed a second consent form before completing the delayed post-test and working memory span tests. All the experimental tasks were administered through E-Prime Professional (version 2.0.10.356). Once all of these tasks were completed, the researcher orally administered a debriefing questionnaire and recorded the participants’ answers on an online form. All participants received monetary compensation for their participation.

5. Analysis and results

5.1. Performance during experimental training conditions

Before addressing the specific research questions, we will present participants’ accuracy on the practice items as this is what was presented in Férnandez (2008), the study that we are extending. On average participants in the EXP condition scored 66.14% ($SD = 16.38$) on the practice items, whereas the IMP condition scored 49.86% ($SD = 8.87$). However, in order to directly compare accuracy scores between the present study and Férnandez, we also consider accuracy scores from the subset of participants who reached criterion as established by Fernández, i.e., four correctly answered items (three subjunctive, one indicative) in a row. Of those who reached criterion in the present study, the EXP condition scored 77.87% ($SD = 15.01$) on average as compared to 60.75% ($SD = 14.13$) for the IMP condition. In both cases, these values are similar to those found for Férnandez’s PI ($M = 78.7, SD = 13.7$) and SI ($M = 67.2, SD = 17.2$) groups, respectively. As for the C+ condition (in which participants did not complete the practice, but received additional exposure and responded to comprehension questions), accuracy scores were near ceiling ($M = 92.90, SD = 7.41$), which confirms that the C+ participants did pay attention to the exposure items.

5.2. RQ 1: Performance on pre-, post- and delayed post-tests

5.2.1. Analysis

In order to answer RQ1, *Do explicit and implicit training conditions differentially affect the development of knowledge of an L2 complex form at immediate and delayed testing?*, a repeated measures ANOVA was conducted to assess between-subjects effects (Condition: EXP, IMP, C+), development over time (Time: pre, post, delayed), and any interactions for accuracy on subjunctive3 (SBJV) items. In addition, development within each condition was assessed by considering confidence intervals (CIs), following Cummings and Finch (2005) and Plonsky (2015) who argue that, for within-group comparisons, mean difference scores (e.g., post-test minus pre-test) with 95% CIs that do not overlap zero can be interpreted as reliable, i.e., statistically significant. Hedges $g$ was also calculated as a measure of the magnitude of any change, and these effects sizes are interpreted according to Plonsky & Oswald’s (2014) field-specific recommendations for within-group comparisons, with .60 indicating a small effect, 1.00 indicating a medium effect, and 1.40 indicating a large effect.

3 Of the L2 target items in the pre-, post- and delayed tests, there were 9 subjunctive and 3 indicative items. The descriptive data suggest that development of subjunctive and indicative items pattern differently, which indicates that they should not be lumped together. However, given the low number of indicative items, we do not have enough power to warrant reporting that data at this time.
5.2.2. Results

Before interpreting the inferential statistics, we consider the overall pattern in accuracy scores in the different conditions (see Figure 2). All three conditions show an increase in accuracy on the immediate post-test followed by a decrease in accuracy at the delayed post-test. Upon further inspection of the CIs around the mean accuracy scores, some noteworthy patterns emerge (see Table 2). Participants in the C+ condition performed below chance at the pre-test, and their performance did not differ from chance on the post- and delayed post-test. A similar pattern is evident for the IMP condition, which starts with below chance accuracy on the pre-test, but does not differ from chance on the post- and delayed post-test. Conversely, the EXP condition did not differ from chance on the pre-test, then performed above chance on the post-test, and then below chance on the delayed post-test.

In order to determine if any patterns in the results were statistically significant, we conducted a repeated measures ANOVA, which revealed a main effect for time, $F(2,63) = 16.63, p < .001, \eta^2_p = .21,$ and a condition by time interaction, $F(4,63) = 3.74, p = .007, \eta^2_p = .11.$ An examination of CIs around the mean differences within each condition (see Table 2) indicates that the accuracy in all three conditions significantly increased from the pre- to post-test, with a small effect size for the C+ ($g = .93$) condition, a medium effect size for the EXP ($g = 1.20$) condition, and a trending to small effect size for the IMP ($g = .51$) condition. In addition, the CIs suggest that the interaction was driven by within-group differences in the EXP condition. More specifically, the decrease in accuracy from post- to delayed post-test in the EXP condition was significant, with a large effect size ($g = 1.38$), whereas this decrease was non-significant with minimal effect sizes in the C+ ($g = .30$) and IMP ($g = .26$) conditions. This would suggest that the C+ and IMP conditions followed the same overall pattern, which differed from that pattern of results in the EXP condition. Overall, changes within the conditions reflect different developmental patterns for participants undergoing different types of training.

![Pre-, post-, and delayed post-test mean accuracy](image)

Figure 2. Mean accuracy for subjunctive items at the pre-, post-, and delayed post-test for experimental and control conditions (C+ = control plus condition, EXP = explicit condition, and IMP = implicit condition).
Table 2
Descriptive statistics of accuracy on post and delayed tests for experimental and control conditions

<table>
<thead>
<tr>
<th>Group</th>
<th>Pre-test M (SD) [95% CI]</th>
<th>Post-test M (SD) [95% CI]</th>
<th>Delayed-test M(SD) [95% CI]</th>
<th>Pre-Post Δ M (SD) [95% CI]</th>
<th>Pre-Post Δ Effect Size for Pre-Post Δ Hedges g</th>
<th>Post-Delayed Δ M (SD) [95% CI]</th>
<th>Post-Delayed Δ Effect Size for Post-Delayed Δ Hedges g</th>
</tr>
</thead>
<tbody>
<tr>
<td>C+ (N = 23)</td>
<td>32.85 (16.23) [25.83, 39.87]</td>
<td>49.57 (18.94) [41.37, 57.76]</td>
<td>44.44 (13.81) [38.47, 50.42]</td>
<td>16.72 (23.13) [6.28, 27.15]</td>
<td>0.93 -5.12 (17.24) [-2.33, 12.58]</td>
<td>0.30</td>
<td></td>
</tr>
<tr>
<td>EXP (N = 19)</td>
<td>46.84 (10.57) [41.75, 51.94]</td>
<td>65.79 (19.09) [56.59, 74.99]</td>
<td>39.77 (17.88) [31.15, 48.39]</td>
<td>18.95 (4.34) [9.82, 28.07]</td>
<td>1.20 -26.02 (20.12) [16.33, 35.72]</td>
<td>1.38</td>
<td></td>
</tr>
<tr>
<td>IMP (N = 24)</td>
<td>41.85 (17.89) [34.30, 49.41]</td>
<td>50.93 (17.31) [43.62, 58.24]</td>
<td>46.44 (16.84) [39.32, 53.55]</td>
<td>9.07 (20.53) [0.40, 17.75]</td>
<td>0.51 -4.49 (25.83) [-6.42, 15.40]</td>
<td>0.26</td>
<td></td>
</tr>
</tbody>
</table>

Note: Δ = Change score calculated as difference between means. CI = confidence Intervals. C+ = control plus condition, EXP = explicit condition, and IMP = implicit condition.
5.3. RQ 2: Individual differences

5.3.1. Analysis

To answer RQ2, Do individual differences (i.e., WM) relate differently to L2 knowledge development of a complex L2 form under different training conditions?, accuracy scores for SBJV items on the post- and delayed post-tests along with the WM composite z-scores were submitted to separate Pearson Product correlational analyses by condition.

5.3.2. Results

As is shown in Table 3, no significant relationships were found at the delayed post-test, but meaningful relationships between L2 development and WM did emerge in all three conditions at the post-test. By condition, we see that this relationship is positive for the C+ (r = .48, p = .03) and EXP (r = .46, p = .055) condition, whereas it is negative for the IMP (r = -.54, p = .007) condition. Although the relationship in the EXP does not quite reach significance, what is important to the present discussion is that when it comes to WM, the C+ and EXP conditions seem to pattern together, but pattern differently from the IMP condition (see Figures 3 and 4).

Table 3
Correlation coefficients of individual difference measures for experimental and control conditions

<table>
<thead>
<tr>
<th>Measure</th>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>C+ (N = 21)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1. Post-test SBJV</td>
<td>-</td>
<td>0.48*</td>
<td>0.48*</td>
</tr>
<tr>
<td>2. Delayed Post-test SBJV</td>
<td>-</td>
<td></td>
<td>0.29</td>
</tr>
<tr>
<td>3. WM Composite Z-Score</td>
<td>-</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>EXP (N = 18)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1. Post-test SBJV</td>
<td>-</td>
<td>0.41</td>
<td>0.46*</td>
</tr>
<tr>
<td>2. Delayed Post-test SBJV</td>
<td>-</td>
<td></td>
<td>-0.05</td>
</tr>
<tr>
<td>3. WM Composite Z-Score</td>
<td>-</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>IMP (N = 23)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1. Post-test SBJV</td>
<td>-</td>
<td>-0.14</td>
<td>-0.54**</td>
</tr>
<tr>
<td>2. Delayed Post-test SBJV</td>
<td>-</td>
<td></td>
<td>0.12</td>
</tr>
<tr>
<td>3. WM Composite Z-Score</td>
<td>-</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note. C+ = control plus condition, EXP = explicit condition, and IMP = implicit condition. SBJV = subjunctive items. WM = working memory. *p < .05, **p < .01
Figure 3. Scatterplots depicting the relationship between the WM composite z-score and mean accuracy for subjunctive items on the post-test for the control plus and explicit conditions.
6. Discussion

The current study extended Fernández (2008) by examining the effect of explicit and implicit conditions on L2 development of a complex form over time as well as the role individual differences may play in this process. With regard to L2 development over time (RQ1), different patterns of development emerged among the three conditions. Specifically, the C+ and IMP conditions significantly improved from below chance performance to performance that did not differ from chance, and they maintained these gains on the delayed post-test. These gains suggest that the participants are beginning to associate the subjunctive with doubt phrases, which may reflect the beginning stages of form-meaning connections, although the participants were never able to correctly distinguish when a phrase should be associated with certainty or doubt. Conversely, the EXP condition showed above chance accuracy at the post-test, but significantly decreased to below chance accuracy at the delayed post-test. In this case, participants started distinguishing when a phrase should be associated with certainty or doubt on the post-test, but failed to do so at the delayed test.

Overall, the results suggest that, for complex forms, explicit conditions may have an immediate advantage over implicit conditions but do not necessarily lead to longer-term development. On the contrary, gains made in more implicit conditions (such as SI practice or meaningful exposure) may be small but can be maintained over time. These results provide evidence for both theoretical predictions regarding the benefit of explicit training on the development of complex forms: on one hand, the EXP condition benefited from EI on the post-test, which would support Hulstijn and Graaff’s (1994) prediction; on the other hand, the implicit condition’s small gains were maintained at the delayed post-test, which would support Reber’s (1989) prediction. More practice items would be needed to determine if more exposure neutralized the effect of EI and/or led to above chance performance in the IMP condition.

As for individual differences (RQ2), different relationships emerged between WM and L2 development under different training conditions at the post-test. Descriptively, the C+ and EXP conditions both showed a medium, positive relationship between WM and L2 development; whereas, the IMP condition showed a medium, negative relationship between WM and L2 development. The WM results for the EXP condition are in line with Tagarelli et al. (2011), but differ from Sanz et al. (2014) and Robinson (2002) who found explicit information to neutralize any benefit of high working memory.

Figure 4. Scatterplot depicting the relationship between the WM composite z-score and mean accuracy for subjunctive items on the post-test for the implicit condition.
Contrary to previous research, which has found WM to be facilitative under more implicit conditions, in the present study WM negatively predicted development in the IMP condition. Most of these studies did not systematically target complex forms, and the one that did (Tagarelli et al., 2016) looked at structural complexity rather than a single complex form, which may explain the different results found here. The WM results from our study also suggest that our IMP training was different from mere exposure. Although, the accuracy scores look similar between the IMP and C+ conditions, the fact that WM is not facilitative under the IMP condition but is under the C+ condition suggests fundamental differences in the underlying processes. Again, more exposure may be needed for these differences to be evident on assessments of L2 development.

Of course, these results must be interpreted in light of the study’s limitations. Most importantly, the study is characterized by low accuracy scores at each test (pre, post and delayed) with the only instance of above chance performance being the EXP condition’s performance on the immediate post-test. One possible explanation for the generally low accuracy scores could be that our participants were not developmentally ready to acquire the subjunctive. It has been argued that beginner to intermediate learners may not have the syntactic foundation necessary to acquire or incorporate the subjunctive into their interlanguage (Collentine, 1995). However, this is unlikely considering the present study used an interpretation task as opposed to a production task, and the structured input format further limited the processing burden of the task.

A more likely explanation for low accuracy is the low number of items during training. Participants were exposed to 30 trials during training, which may not be enough to lead to above chance performance on assessments of a complex form without explicit information, which indeed is a critique of previous empirical comparisons of explicit and implicit conditions. Further support for this explanation can be seen when the scores of the present study are compared to Farley (2001a), who examined the subjunctive under PI (similar to the current EXP condition) versus meaningful output instruction. When compared, the pre-test accuracy scores for our EXP condition are similar to those found in Farley’s (2001a) PI condition ($M = 39.8\%$). Nevertheless, participants in Farley’s (2001a) PI condition evidenced improved accuracy at an immediate and delayed post-test (1 day: $M = 85.3\%$, 1 month: $M = 83.1\%$). It is important to note that Farley’s participants were exposed to eight PI activities, consisting of approximately 99 subjunctive tokens, over the course of two class periods for a total of 90 minutes of practice; whereas, in the present study participants were exposed to a single task consisting of 30 subjunctive tokens that lasted around 10 minutes. There is a clear disparity in the amount of exposure our participants received compared to Farley (2001a), and it is expected that with more subjunctive tokens both experimental conditions would evidence above chance gains, at least on the post-test. Future research should include more items in order to increase exposure to the form and thereby increase the potential for development.

7. Conclusion

Exploring the potential interaction between training condition, linguistic complexity, and individual differences has recently moved to the forefront of research (Housen & Simoens, 2016; Tagarelli et al., 2016). This question is of theoretical importance, but it is also of ever increasing importance in applied linguistics as more and more L2 acquisition is taking place in foreign language contexts devoid of target language input outside of the classroom. Thus, tailoring classroom exposure to the needs of specific learners and the complexity of the target form are issues teachers wrestle with on a daily basis. The present study contributes to this body of literature by examining the role of explicit and implicit training and working memory on the development of a complex linguistic form.

Specifically, the present study improved on methodological limitations of previous studies by balancing the pre-practice exposure across conditions and providing the same meaningful practice in both experimental conditions. Based on the results, we conclude that, for complex L2 forms, there seems to be an initial advantage for explicit information reported, as seen generally in previous research. However, the correlational results from the present study indicate that the participants in the EXP condition may have relied on WM. In general, reliance on strategies available to short-term working memory may have contributed to the fact that initial advantages for explicit information may be limited to the immediate assessments. In contrast, for the IMP condition, WM was negatively associated with L2 gains yet the gains were maintained over time, even though learners had not yet achieved the full ability to interpret the complex L2 form. Thus, implicit conditions may lead to retention that are not
dependent on the engagement of working memory. Overall, the results suggest that type of exposure differentially influences the initial stages of L2 development, and that the role of WM also varies by type of exposure. Future research and replications are needed to further substantiate the robustness and generalizability of the present findings. Data collection is ongoing to have the adequate statistical power to replicate Fernández (2008) along with the extension reported here. In addition, we are collecting confidence ratings and subjective measures to delve deeper into the type of knowledge that is developed under the different conditions.

References


Williams, Jessica, & Evans, Jacqueline. (1998). What kind of focus and on which forms? In Catherine Doughty & Jessica Williams (Eds.), *Focus on Form in Classroom Second Language Acquisition* (pp. 139-155). Newyork: Cambridge University Press.