Arizona Working Papers in SLA & Teaching, 19, 56-81 (2012)

THE MAZE TASK: TRAINING METHODS FOR SECOND LANGUAGE LEARNING

Elizabeth Enkin
University of Arizona

ABSTRACT

The maze task was created for psycholinguistic experimental testing (Forster et al., 2009). However, this paper explores the merits of this task as a language training program for beginning Spanish learners. The attributes of providing ample comprehensible input and immediate corrective feedback allow the maze task to be considered as a potential supplemental pedagogical tool. Moreover, transfer effects to implicit and explicit measures as well as students' perception of such a task are examined.

INTRODUCTION

The maze task is a psycholinguistic technique used in experimental testing that records reaction times as subjects read (and comprehend) sentences. The task asks subjects to "weave" their way through a sentence word by word by choosing the correct grammatical alternative from two choices (Forster, Guerrera, & Elliot, 2009). The current study's main question asks if the maze task can be applied to a teaching program. In other words, could training on particular sentence types using the maze task help late L2 learners to better their foreign language performance? If the maze task does in fact yield training effects and learning benefits, is it a task that is enjoyable for students, and why? Thus, it is the intention of this paper to provide a psycholinguistic framework from which to draw pedagogical implications.

The foundation for this paper rests on the implicit and explicit learning dichotomy and explores the merits of integrating both types of instruction within a late L2 learning curriculum. Explicit learning is associated with selectivity, which presupposes a deductive, concept-driven mode of learning; on the other hand, implicit learning is associated with unselectivity and assumes an inductive, data-driven mode of learning (Gasparini, 2004; N. Ellis, 1994). One of the main questions in second language learning is what type of instruction is best for L2 acquisition. Implicit learning is the retrieval and use of memories that have been formed without conscious awareness, whereas grammar rules and guided instruction are illustrative of explicit learning.

Similarly, implicit *knowledge* is the intuitive understanding of the manner in which a language works; whereas explicit *knowledge* is conscious awareness of the grammatical rules of a language (R. Ellis, 2009a). Within the constructs of both connectionist and generative accounts of linguistic

competence, there is general agreement that linguistic knowledge is primarily comprised of intuitive and tacit knowledge (N. Ellis, 2005; R. Ellis, 1993). It may be the case, however, that adult L2 learners necessitate explicit knowledge due to the role of the L1 and its transfer effects (DeKeyser & Juffs, 2005; R. Ellis, 1993). The question then becomes what is the best mixture of implicit and explicit learning for late learners in order to build the implicit knowledge base.

Implicit and Explicit Learning and Form-Focused Instruction

The interface between implicit learning yielding explicit knowledge and explicit learning yielding implicit knowledge is not certain, and is therefore difficult to assert which learning mode would best lead to an implicit knowledge base (N. Ellis, 2005; Hulstijn & DeGraaf, 1994). Although implicit knowledge may be necessary for eventual success in L2 learning, at least some form of explicit instruction may be necessary with late learners in order to facilitate acquisition. Explicit instruction, also known as form-focused instruction (FFI), may actively aid in drawing metalinguistic attention to an L2 structure, which is more helpful than having students merely notice a form (R. Ellis, 2002). Moreover, due to classroom environment constraints such as time and limited language input and use, a curriculum should include both explicit and implicit instruction, especially if ultimate success could depend on the implicit knowledge base.

One type of methodology that is based around both explicit and implicit instruction is derived from Input Processing (IP), which is a comprehension-based theory developed by Van Patten (2004). From IP theory, Van Patten (2004) has illustrated a type of instruction, namely Processing Instruction (PI), which is a type of FFI. PI attempts to aid learners in developing richer intake from input by way of having learners engage in structured input activities that aim to push learners into the correct strategies of processing for meaning (Wong, 2004). Structured input caters to implicit learning insofar as it requires students to arrive at their own restructuring processes while processing the given input for meaning (R. Ellis, 2009a). Once linguistic competence is acquired through this method, the same knowledge base is drawn upon during comprehension and production. The model of PI is found in the figure below.

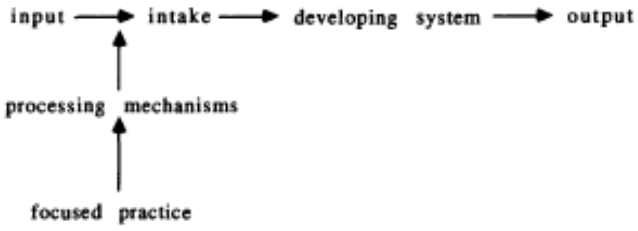

Figure 1: Processing Instruction in Foreign Language Teaching (Van Patten & Cadierno, 1993:46)

Van Patten (2004) suggests that linguistic competence in an L2 is acquired primarily through implicit learning although some explicit instruction is necessary. To this end, PI merges implicit and explicit instruction by first raising metalinguistic awareness through two stages of processing – explicit explanations followed by implicit instruction through practice with structured activities (Van Patten & Cadierno, 1993). PI supports the role of formal instruction in creating representations in explicit (declarative) memory. However, it also emphasizes the significance of practice and its role in converting what was learned into implicit (procedural) representations. Thus, PI has the potential of incorporating the best of both worlds – it assumes that both implicit and explicit instruction have a place in the L2 classroom.

The maze task is also based around comprehensible input, and when used as a training instrument, aims to assist in altering processing strategies when needed. By using the maze task as an implicit type of instruction, it can be seen as a compliment to more formal, explicit teaching, which occurs within the classroom construct. In other words, the maze task borrows from PI in the sense that it is meant to be used in conjunction with explicit instruction so as to better assist in ultimate L2 attainment. It is through the process of correctly building a sentence, word by word (where each correct word is presented simultaneously to the learner with an incorrect alternative), where any strategy that is incorrect, can be altered. It is also important to note that although the task is mainly implicit in nature, it does contain a minimally invasive element of explicit instruction. It provides immediate feedback in the precise location where an error occurs, which aids in strengthening the new neural networks made during class instruction. In this way, formal instruction is reinforced by way of implicit practice. This then leads to converting explicit knowledge into implicit representations, which is necessary for eventual fluency.

To test whether the maze task can aid L2 learning more generally, post-tests were given to participants in order to investigate generalizing effects for language learning as a whole. One type of computerized post-test that measures explicit knowledge is an unspeeded grammaticality judgment task (R. Ellis, 2009b; Loewen, 2009). This post-test is an integral component of the project since it assists in exploring whether implicit learning through the maze is effective in building the explicit linguistic knowledge base as well. In other words, if there are benefits seen on this explicit measure, this would lend support for the hypothesis that implicit maze training may assist in constructing the explicit knowledge base as well. Results from this task can create a clearer picture of how the maze task is affecting the two linguistic knowledge bases.

The aforementioned task does examine the amount of explicit knowledge students have; however, only the comprehension domain is

investigated. Therefore it was necessary to develop a pre-test/post-test design in which production was measured. This would allow for an evaluation as to whether benefits from the maze task, a comprehension based task, can generalize to a production based activity. The premise for the maze task is similar to Van Patten's (2004) viewpoint on comprehensible input being the building blocks of the knowledge base from which comprehension *and* production is drawn. However, the literature is mixed regarding whether comprehension training can generalize to production (Van Patten, 2004; Morgan-Short & Bowden, 2006). Thus, by using a pre-test/post-test design where improvement is monitored on a target structure (from the maze task), generalizing effects, and thus learning from maze task training, can be determined.

The Influence of the L1: "Hard" and "Easy" Sentence Types

To test the effectiveness of maze task training, it is necessary to assess how more complex (harder) syntactic constructions are learned versus how simpler (easier) structures are handled since learning an L2 involves both types of sentences. Thompson, Shapiro, and Sobecks (2003) investigated whether or not training of syntactically complex filler-gap sentences could generalize to performance on syntactically less complex filler-gap sentences. Their results indicated that robust generalization effects for less complex sentences surfaced for individuals who were trained on the more complex structures. These results also illustrate the generalizing effects of implicit training since subjects carried over their skills from a comprehension task to a production test, which is something the maze task aspires to do as well. Moreover, this study reveals an important facet of implicit learning—in the face of the complex rules that students must learn when acquiring a L2, it may be better for them to use an implicit mode of learning. Succinctly stated, rule learning may work better for simple constructions, whereas more complex L2 constructions may necessitate additional implicit instruction such as training via the maze task.

When thinking about more complex (harder) constructions as compared to simple (easier) constructions, the impact of the L1 becomes an important aspect to consider. Structures with the same word order between the L1 and the L2 generally pose less of a problem during L2 acquisition. Thus, they are referred to as easier, more simple constructions (or matched structures) since they would not include a shift of parameter values from the L1 (DeKeyser & Juffs, 2005). To illustrate the matter further, in an experiment by Tokowicz and MacWhinney (2005), event-related potentials (ERPs) were used to investigate whether or not L2 learners of Spanish were sensitive to matched and mismatched structures between the L1 and L2. The results indicated that learners displayed a P600 effect (indicating awareness of grammatical violations) with structures unique to the L2 (such as determiner gender marking in Spanish vs. none in English) as well as those structures that were similar in both languages. Nevertheless, learners were not sensitive to constructions that differed in the L1 and L2 (that is, those structures that

existed in both languages but contained contradictory properties in some surface forms). These results emphasize that in order for an implicit knowledge base to be formed, a learning tool must be able to change the processing strategies of harder, more complex constructions, or those that show a mismatch between the L1 and L2.

It does seem possible, however, to overcome L1 parameter settings as seen through training. Nitschke, Kidd, and Serratrice (2009) examined L1 transfer effects on L2 sentence comprehension by manipulating preference of subject and object relative clauses (SRCs and ORCs). By having a prime—target—post-test design, the results of their study indicated that training a mismatched condition through a prime-target construct yielded long term structural priming on a post-test. That is, after the prime-target phase, L2 learners were more likely to indicate in the post-test that the sentence was an ORC interpretation even when the word order was a mismatch from their L1. Thus, these results act as reinforcement for the idea that maze task training can yield a successful long-term parameter shift for harder, more complex constructions.

Sociocultural Theory and the Maze Task

Another aspect of the maze task revolves around its fit into sociocultural theory (SCT) insofar as it is a pedagogical instrument. In regards to education within sociocultural theory, Vygotksy (1978) outlined the significance of the zone of proximal development (ZPD). The ZPD can be defined as the difference between what students can do with the help of an expert (the teacher) and what they can do on their own without help (Vygotsky, 1978). The maze task strives to keep students in the ZPD by providing immediate "expert" feedback so as to help mediate their thinking process. With respect to Van Patten (2004) and PI, this feature assists in making the input more "structured" as well.

By providing comprehensible input of varying difficulty level (hard and easy), the maze task also aims to situate students within a learning environment that is level-appropriate, yet still challenging. This type of design keeps students interested and motivated on the task at hand while keeping anxiety low, which is important for language learning to take place (Krashen, 1994). In this way, maze task training can be seen as a useful supplementary homework activity for L2 learning. Furthermore, experimentally testing these two different sentence types will assist in determining the most optimal training method. It is therefore the purpose of this paper to combine psycholinguistic research with pedagogical theories so as to bridge the gap between these two areas in the field of second language acquisition.

In the following pages, three experiments are devoted to empirically testing how implicit maze training fares with foreign language learners. More specifically, the benefits of training with more complex vs. simpler sentence types are analyzed and discussed. Lastly, results on post-tests measuring explicit knowledge will offer extensions to further benefits of maze training.

Experiment 1

Research Questions

1. Does training aid in more automatic completion of new sentences, and if yes which type (training with simple or more complex sentences)?
2. Is the maze task a good teaching tool insofar as a learning/training instrument, and if yes, which type of training is best?
3. Does implicit training in comprehension (by way of the maze) generalize to a production task measuring more of explicit knowledge, such as a paper-and-pencil worksheet?
4. What are students' general opinions to such a task as revealed through a questionnaire?

Method

Participants. Forty-four (44) subjects enrolled in Spanish 102 for course credit participated in the maze task. An additional forty-one (41) subjects also enrolled in Spanish 102 were given course credit for their participation in the paper pre-test/post-test. In this and all subsequent experiments, participants were native speakers of English. The forty-four (44) maze task subjects were randomly assigned into one of two training groups, which was either "English" (Eng) or "Spanish" (Spa) groups. The only difference between these groups was the types of sentences they completed during training sessions. The Eng group received English-similar (ES) sentences that contained word order and lexical items similar to their native language (English). On the other hand, the Spa group received more complex, Spanish-specific (SS) sentences that contained word order and lexical items that were specific to their L2 (Spanish).

Materials and Design

There were two different types of sessions that were involved – training sessions and a test session. Three training sessions were completed by both training groups over a three week period with a frequency of one per week. Twenty (20) sentences composed each training session, and fifteen (15) of these contained target structures (ES or SS depending on the training group - Spa or Eng). The sentences were the same from session to session, but the location of words was randomized so that memory of correct word location would not be a confounding variable. During the training sessions, subjects were asked to try the sentence again if they made a mistake. The location of the mistake in the sentence was pointed out immediately so that students could see where they had made an error.

A final test session was administered on the fourth week, which contained all new sentences, but of both English-similar (ES) and Spanish-specific (SS) types. There were a total of thirty-two (32) sentences, twenty-eight (28) experimental sentences, and subjects were not able to try the sentence again if they made a mistake. The feature of immediate feedback was still present.

There were four sentence types used in this experiment, which were the target structures used for both training and test sessions. They are as follows: subject relative clauses (SRCs), object relative clauses (ORCs), direct object pronouns (DO), and the verbs "to be" (TB). SRCs were added to create more variety within the sentences, and the same sentences were used in both training groups. On the other hand, ORCs, DOs, and TB constructions all had an "easier" version and a "harder" version. As explained above, the harder, more complex versions were called the "SS" types because they had a "Spanish-specific" order. In other words, these sentences contained L2 structures that are not found in the L1. On the contrary, the easier, more simple sentences were called the "ES" types since the critical structures contained "English-similar" constructions. Simply stated, the sentences were similar to L1 constructions. Each training group received these four sentence types. The only difference was that the Eng group received the easier versions of these sentences (ES) while the Spa group received the harder versions of these constructions.

With the ORCs, the easier constructions contained an overt subject after the relativizer whereas the harder sentences contained a pro-drop construction. In the DO condition, the clitic appeared in the same location as it would in an English sentence (post-verbal), or it was raised (unlike in English) in the harder constructions. Finally, the easier constructions contained the primary "to be" verb, *ser*, which expresses permanency (and is more readily assimilated into its English translation). Constructions focused on specific uses of the verb, such as describing professions and expressing time. The harder constructions contained its counterpart, *estar*, which expresses temporary states, and does not exist in English. These sentences focused on uses such as describing emotion and location. The table below illustrates sample sentence types.

Sentence Types	***English-similar Sentences***	***Spanish-specific Sentences***
Object Relative Clauses	*El vino tinto que el hombre* pone en la mesa es de alta calidad. (*The red wine that the man* puts on the table is of high quality.)	*Los pantalones que Ø tienes* son muy populares pero no son baratos. (*The pants that you have* are very popular but are not cheap.)
To Be	Trabajo en una oficina grande porque *soy* abogada. (I work in a big office because *I am* a lawyer.)	No quiero salir porque *estoy* triste esta noche. (I do not want to go out because *I am* sad tonight.)

Direct Object Pronouns	El libro de misterio es muy interesante y *voy a leerlo* pronto. (The mystery book is very interesting and *I am going to read it* soon.)	Escribí las cartas ayer y *las quiero enviar* hoy. (I wrote the letters yesterday and *I want to send them* today.)

Table 1: Sentence Types

A paper-and-pencil pre-test and post-test was administered to students who had gone through the maze task series and also to two other Spanish 102 classes that did not have maze training. This was done so that the benefits of the maze task could be established more clearly. The pre-test values were derived from scores on a subsection from the students' exam #1 (taken before maze training), which tested their ability to distinguish between *ser* and *estar* in a fill-in-the-blank format. This covered the "to be" constructions from the maze. The post-test was also a fill-in the blank worksheet on *ser* vs. *estar* where students had to choose and correctly conjugate the verb. This was essentially the same task as the section from exam #1 with the exception being that these were different sentences. As with the post-training maze session, the post-test *ser* vs. *estar* worksheet contained equal amounts of ES and SS sentence types (5 and 5) for a total of 10 sentences all together. The pre-test/post-test items were considered holistically for both groups combined in order to investigate whether maze training itself, regardless of type, had an effect on the explicit knowledge base (and production skills).

Lastly, participants were asked to fill out a survey about the maze task. In general, questions asked for feedback on the likeability and usefulness of the task. Subjects were asked to rate each question on a scale from 5 to 1, with 5 being "strongly agree" while 1 was "strongly disagree" (see appendix for complete questionnaire).

Procedure

The experiment was run using DMDX software, which was developed by J.C. Forster and K.I. Forster at the University of Arizona (Forster & Forster, 2003). The items were presented in black letters on a white background. Every item, each making up a sentence, consisted of a series of frames. For example, the first frame of each set of items making up a sentence would look like as follows: [La …]. Each subsequent frame contained two words side by side, where one was the correct next word in the sentence, while the other was grammatically and semantically incorrect. Incorrect alternatives were not the same part of speech as the correct choice, which eliminated this issue as a potential confound.

Because training sessions contained the same sentences and incorrect alternatives for each training group, the incorrect alternatives appeared on random sides (left or right) from session to session. This was done so that

students could not memorize the correct alternative's position on the screen. Sentences were presented in a randomized order for each subject for each session. Every maze session was sent via email as a link, and once students clicked on a link, DMDX (software used in psycholinguistic experimental testing) would automatically install on their PC for the duration of the task. Subjects completed each session in one sitting and only one time. They had a full week to complete each session so as to allow them to do each one at their convenience.

Participants were instructed to choose the correct word in each frame as quickly and as accurately as possible by pushing the corresponding left or right button. If the word was correctly selected, the next frame was displayed immediately. If the incorrect alternative was selected, an error message was displayed. If an error occurred in the training sessions, subjects were given the choice to try the sentence again by pushing the corresponding key. In the test session, however, they were not given this choice, and when an error occurred, the program moved onto the next item (the start of a new sentence). If the participant made the correct choice throughout the frames for an item, the final frame was followed by a "CORRECT" message. Subsequently, the beginning of the next item would appear. Thus, in table 1 above, each word of each sentence represents its own item (and frame) with a corresponding incorrect alternative next to it.

Paper tests were given to students upon the completion of all maze tasks. The paper and pencil pre-test and post-test was administered a week prior and a week after all maze sessions were completed, respectively. An outside class not having taken the maze task training series also participated in these paper and pencil tasks, and they took the tests at the same time in the semester as subjects undergoing maze training. These tests were administered during class time as in-class work. Lastly, an eleven (11) question questionnaire asking students to rate their experience with the maze task was also administered in class on the same day as the post-test.

Results and Discussion

All analyses were carried out using linear mixed effects modeling. Thus, the analyses involved fitting linear mixed effects models (LMERs) to the data, which was done using the lmer function from the lme4 package in R (Baayen, 2008a, 2008b; Baayen, Davidson, & Bates, 2008; Pinheiro & Bates, 2000; R Development Core Team, 2009). Unlike traditional analyses carried out through ANOVAs, the method using LMERs allows for two crossed random effects (subjects and items). The software analyzes the data for each individual trial, without needing to aggregate over items or subjects, and then can arrive at the best fitting linear model with subject and items as random effects. The *p*-values for the effects were generated by Markov Chain Monte Carlo (MCMC) simulation, which uses 10,000 iterations (Baayen et al., 2008).

Prior to all analyses, the raw reaction times (RTs) as well as error rates, which were the dependent variables, were log converted in order for the data to reflect more of a normal distribution. All trials where an error occurred were discarded. In addition, trials that were never seen due to an error were all

discarded. This occurred if the subject would error out of a sentence (i.e. when a subject made a mistake and was automatically redirected to the next sentence) thereby not ever seeing the rest of the sentence. Subject relative clauses (SRCs) were also removed from the analysis since they acted as fillers only. Lastly, RTs were trimmed so that those under 300 and over 5000 were not included in the analysis.

All items were considered holistically, and therefore no particular region was specified for analysis. This was done because the central question being investigated was whether or not the maze task could be used as a training program (meaning, the entire sentence). It was predicted that training with more difficult constructions would yield greater benefits as seen on the test session containing both sentence types. This would complement the results from aphasia research demonstrating that training in the harder to easier direction yields stronger gains (Thompson, Shapiro, & Sobecks, 2003). Thus, it was hypothesized that more difficult training would show learning outcomes insofar as sentence completion RTs for both types of sentences.

For Experiments 1-3, there were two factors that were considered in the analysis. The first was the effect of Training Group (Trggrp), with the levels TrggrpEng (for easier, L1-like word order sentences) and TrggrpSpa (for harder, L2-specific word order sentences). The second factor was Sentence Type (Stype), with the levels StypeES (for "English-similar" word order sentences) and StypeSS (for "Spanish-specific" word order sentences).

Using reaction times as the dependent variable, the first result of note showed that although there was no main effect of stypeSS (t = 0.56, p >.05) nor of TrggrpSpa (t = 1.58, p >.05), the critical interaction of StypeSS:TrggrppSpa was significant (t = 2.61, p = .01). This demonstrates that students trained on harder constructions (those that had specific L2 word order) yielded significantly comparable, and faster, reaction times on both hard and easy constructions as compared to the easier training group. This suggests that the more complex training assisted students in understanding the distinction between the two sentence types. The mean reaction times can be found in the figure below.

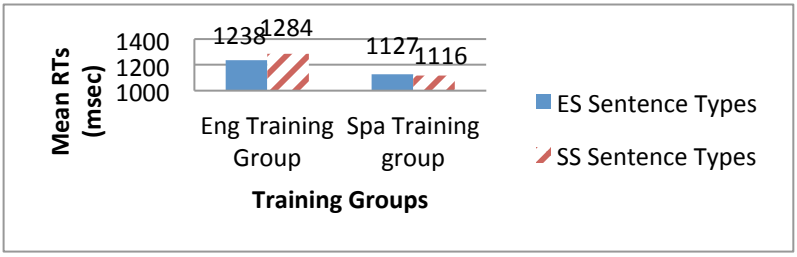

Figure 2: Experiment 1 Mean Reaction Times

This result indicated that the Spa training group completed SS sentences faster than ES sentences, and moreover, their reaction times on both

types of sentences (ES and SS) were comparable. With respect to Vygotsky (1978) and his theory of the ZPD, it could be that harder sentences assist in keeping students in the "zone" more so than easier constructions. Moreover, this effect echoes past findings with aphasia research demonstrating that harder constructions will generalize to performance on easier sentences, but not the reverse (Thompson, Sobecks, & Shapiro, 2003). Although the result in this experiment deals with sentence types unlike constructions from the aphasia research, it stands to reason that a similar mechanism may be at play, and is therefore worth noting. That is, when participants are trained on more difficult constructions, it is then easier for them to understand the different usages between complex and simpler sentences.

Next an analysis was carried out to examine the maze task's effectiveness with regard to its generalizing potential to a paper-and-pencil production pre-test/post-test. Two main factors were considered. The first factor was Paper Test (PTest) with levels Pre-test and Post-test. The second factor was Class with the levels MClass and NMClass for the classes that underwent maze training and ones that had no maze training, respectively. By comparing two other Spanish 102 classes that did not go through maze training (NMClass) with the two classes that did (MClass), the effect of factor Paper Tests (PTest) by Class was analyzed. A significant interaction of Pre-test: Mclass (t = 2.63, p =.01), revealed that as compared to the control classes, students that completed the series of maze tasks showed significant improvement on their post-test as compared to their pre-test scores. This result suggests that by undergoing maze training, students show a benefit on a production measure.

Lastly, the questionnaire revealed an average score of 4.33/5 on all questions. Top scoring questions included: 1) the maze task is a great supplement to online workbooks/ more enjoyable (perhaps due to its interactive nature) (question 3), 2) it would be a good addition to the Spanish curriculum (question 11), 3) it could help with other languages (question 7), 4) it was overall extremely helpful (question 2), and 5) it helped students learn Spanish (question 4) and they think it could help others (question 6), and lastly 6) students wanted to try sentences again if they made an error (question 5). Students also thought that maze task training could benefit other modalities of assessment (such as papers and tests) (question 8).

Experiment 2

Research Questions

The same research questions as Experiment 1 were investigated, and thus Experiment 2 served as a replication. In addition, a further computerized post-test was included so as to further test the generalizing effects of the maze task. This task was an unspeeded grammaticality judgment task, which is a task that falls in the comprehension domain, and measures explicit knowledge (R. Ellis, 2009b; Loewen, 2009). The error rates were analyzed.

Method

Participants. Twenty-one (21) subjects enrolled in Spanish 102 participated for course credit. Training groups were the same as in Experiment 1.

Materials and Design. The materials and design were identical to Experiment 1, except with some differences in the post-test design. Rather than using a sub section of an exam as a pre-test, two versions of a *ser* vs. *estar* fill-in-the-blank worksheet were created. A counterbalanced design was implemented in that half of the subjects received version A as their pre-test whereas the other half received version B as their pre-test. For the post-test, subjects received the version they had not completed yet. The two versions were of the same exact skill level (suitable for Spanish 102), with each one containing 5 ES types and 5 SS types. The only difference between versions was lexical items used, which were all of appropriate level. However, only the class undertaking the maze task series completed this pre-test and post-test due to lack of availability of another class.

In addition, an unspeeded grammaticality judgment task was used as another post-test measure evaluating explicit knowledge. Both types of sentences (ES and SS) were included so as to observe generalizing effects of the maze task. Again, only the class going through the maze task series took this post-test. The link for this post-test was sent through email as well.

Procedure. The same procedure was used as in Experiment 1.

Results and Discussion

The same statistical analysis software (R) was used for data analysis, and the same factors were considered. Using RTs as the dependent variable, it was predicted that Experiment 1's finding would be replicated. Indeed, this critical interaction was significant, thus displaying that the interaction StypeSS:TrggrpSpa (t = 2.33, p =.02) was significant once again. This indicated that reaction times on the test session were replicated so that subjects receiving training with harder constructions completed both sentence types (ES and SS) just as quickly. The mean reaction times can be found in the figure below.

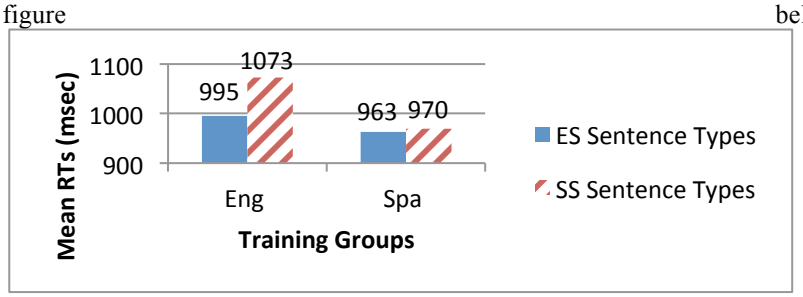

Figure 3: Experiment 2 Mean Reaction Times

In regards to the unspeeded grammaticality judgment task, all incorrect trials were discarded. Using error rates on the grammatical items as the dependent variable, there was a main effect of TrggrpSpa, (t = 2.10, p <.05), indicating that the subject group trained on the harder constructions

(SS) made significantly fewer errors overall, but only on easier constructions (ES). This could mean that though this group's knowledge may be the same as the "English-similar" (Eng) training group, their *performance* becomes better on the English-similar sentences. This could imply that training on harder constructions in the maze task actually helps with performance on easier constructions (but not on the harder ones with which they were trained). The figure below shows the mean error rates for this task.

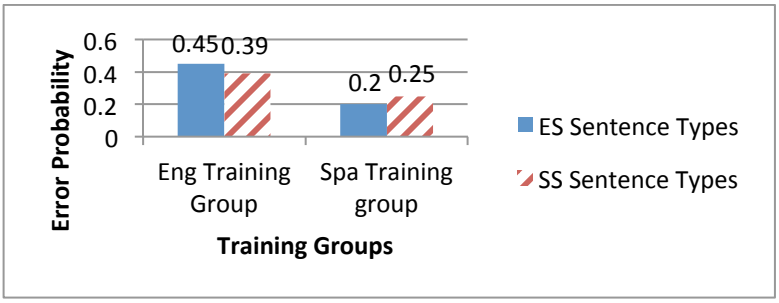

Figure 4: Experiment 2 Grammaticality Judgment Task Error Rates

When the production paper-and-pencil task was analyzed, there was significant improvement found from the pre-test to post-test by a related sample one-tailed within-subjects t-test, (t [20] = 1.95, p=.03). This indicates that after undergoing maze training, subjects significantly improved on a measure of production where a structure from the maze was tested. Thus, completing maze task training itself can have significant benefits on production skills and the explicit knowledge domain when content is similar.

Lastly, the results of the questionnaire were replicated with an average of 4.31/5. The top scoring questions included that it was a fun and helpful task (questions 1 and 2, respectively), that it was an enjoyable supplement to online workbooks (question 3), that they wanted to try the sentence again when an error occurred (question 5), that it could be helpful for other languages (question 7), and that it would be a good addition to a foreign language curriculum (question 11).

Experiment 3

Research Questions

The same research questions from Experiments 1 and 2 applied here, but with one additional measure taken. In Experiment 3, more experimental items were added in order to further examine if the training effect was genuine. In other words, it could be the case that there is no generalizing effect of harder constructions to easier constructions because students may just have ES type constructions set as their default. Thus, increasing the amount of experimental items would provide additional insight into the maze task effectiveness, specifically with respect to the more difficult training type.

Method

Participants. There were twenty-four (24) subjects enrolled in Spanish 102 that participated for course credit. Training groups were identical to those from Experiments 1 and 2.

Materials and Design. The materials were identical to those of Experiment 1 and 2, but the final maze test session now included thirty-two (32) experimental sentences rather than twenty-eight (28).

Procedure. The procedure was the same as in Experiment 2. The only difference was that rather than having only 10 fill-in-the-blank sentences for the *ser* vs. *estar* worksheet, another 10 sentences (5 ES types and 5 SS types) were added so as to leave more room for potential improvement. This totaled 20 sentences rather than 10.

Results and Discussion

Once again, the data was analyzed using LMERs in the R statistical analysis software, and all factors remained the same. All low and high cutoffs remained the same as in Experiments 1 and 2. Furthermore, incorrect trials were discarded once again when considering RTs as the dependent variable. When looking only at Experiment 3, now with additional items as compared to the previous two experiments, the training effect as revealed on the final test session showed a significant main effect of TrggrpSpa ($t = 1.92$, $p < .05$). This indicated that this group yielded significantly faster reaction times on both types of sentences. Furthermore, taking error rate as the dependent variable yielded a significant main effect of TrggrpSpa, ($t = 2.0$, $p = .046$), which illustrated that there was a training effect for this group in regards to error rate as well. Together, these results suggest that there is a genuine training effect occurring for the group receiving training with more complex constructions (Spa). That is, due to the significant main effects, it cannot be the case that the ES sentence types are simply easier (or default types) for all students.

In Experiment 3, the critical interaction of StypeSS:TrggrpSpa did not reach significance ($p > .05$). However, in order to heighten the power of the experiment due to the low number of subjects, participants from Experiments 2 and 3 were combined for a total of 45 subjects. Using RTs as the dependent variable, the interaction StypeSS:TrggrpSpa reached significance again, ($t = 2.53$, $p = .01$). This finding replicated the result that training with more complex sentences aids students in understanding the distinct usages between the two sentence types. The mean reaction times combining Experiment 1 & 2 can be found in the figure below.

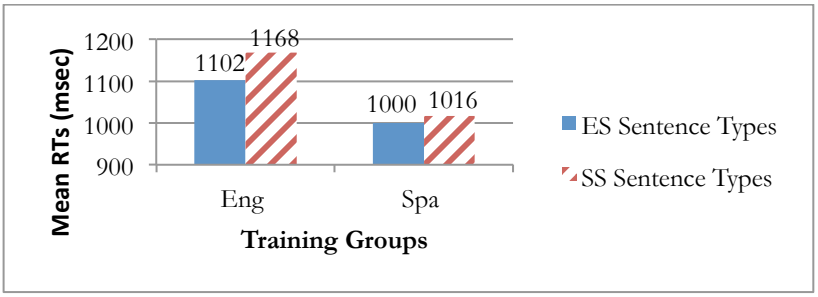

Figure 5: Experiments 2 & 3 Mean Reaction Times

When it comes to the unspeeded grammaticality judgment task, all incorrect trials were discarded again. Taking error rates as the dependent variable for all grammatical items, there was a significant main effect of TrggrpSpa (t = 2.03, p < .05) signifying that the Spa training group made significantly fewer errors overall. There was also a significant interaction of StypeSS:TrggrpSpa, (t = 2.11, p = .03). This signified that the difference in performance between SS types and ES types was significantly greater than the difference exhibited by the Eng training group.

To make certain that this result illustrated a true effect from maze training, it was necessary to see if the interaction would reach significance using N (neither sentence type) as the baseline condition (where N represents an additional level of the factor "Sentence Type"). This would illustrate that not only is performance on SS sentence types as compared to ES sentence types significant, but also that the training group receiving more complex structures did better on SS sentence types as compared to N types as well. Importantly, this displays the genuine carryover effect from maze training that is occurring. Under these conditions, the same interaction did reach significance (t = 2.58, p =.01). The following figure illustrates the mean error rates.

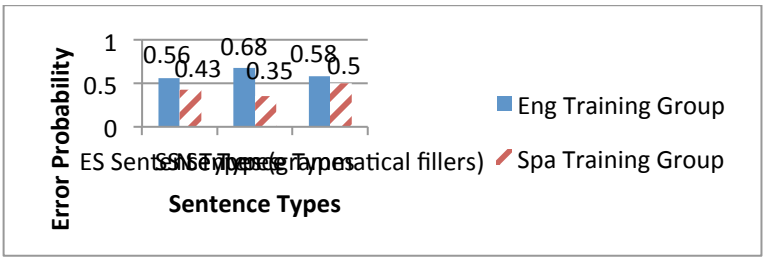

Figure 6: Experiment 3 Grammaticality Judgment Task Error Rates

In regards to the paper-and-pencil task, taking Experiment 3 only, a related sample one-tailed within-subjects t-test revealed a strong trend, (t [11] = 1.69. p =.06), but this was *only* for the Spa (more complex) training group.

This indicated that only the Spa training group significantly improved from pre-test to post-test. When combining the subjects from Experiments 2 and 3 for a total of 45 participants, a related sample two-tailed within-subjects t-test revealed a significant effect of TrggrpSpa (t [21] = 2.13, p = .04). This analysis indicated that training with harder structures yields significant improvement on a measure of explicit knowledge as well as a test of production. This suggests that the effect of more complex training through the maze task can generalize to other tasks. Moreover, completing this type of maze task training can have a positive effect on the explicit knowledge base as well as on production skills.

Lastly, the results from the questionnaire were replicated once again in regards to opinions to such a task. The average of 11 questions was 4.3. Top scoring questions revealed that the maze task was a fun and helpful activity that was more engaging than traditional online workbook activities (questions 1 through 3). Once again, students expressed that the maze task would be a good addition to a basic foreign language program (question 11). They also wanted to try the sentence again if they made an error (question 5). Overall, students communicated that the maze task was a helpful activity, and could be helpful for learning other languages other than Spanish (questions 6 and 7, respectively).

General Discussion

The maze task has exemplified a type of training program that could have potential gains
when it comes to computer assisted language learning, specifically for beginning learners. The task was extremely well received by the students involved in the study (as seen through the questionnaire results), which suggests heightened interest in incorporating such an activity into a foreign language curriculum. Supporting past research, it was found overall that more complex training (Spa training group) yielded higher gains as compared to less complex training (Eng training group). Moreover, results from Experiment 3 alone suggested that because the Spa training group performed better overall, it could not be the case that the ES sentences were just the "default" types. This finding lends further support regarding the reliability of results – that training with more complex structures yields significant benefits for learners.

The maze task assists students with more automatic responses to sentence comprehension when the input contains challenging sentence constructions. The question of automaticity being developed through implicit training programs is one that has been written about extensively (DeKeyser, 1997; N. Ellis, 2005; Robinson, 1997). Through the experiments in the present paper, it has been shown that implicit training with more complex sentences aids in more automatic completion of L2 sentences of all different types. Thus, the implicit knowledge base, the foundational element in developing L2 fluidity, is being constructed through such a method. Connections made during class instruction are strengthened through this type of implicit training.

Both training groups also showed benefits from the maze as demonstrated through a pre-test and post-test design. In Experiment 1, both groups showed significant improvement from pre-test to post-test as compared to other classes not undergoing maze training. This demonstrated that the maze task's training within the comprehension domain has the capacity to generalize to the production domain when content is similar. Moreover, training with the maze generalizes to an explicit type of task. Experiment 2 also showed this improvement when both training groups were analyzed together again. Experiment 3, however, only yielded improvement for the Spa training group. When combined with subjects from Experiment 2, this improvement was significant. Thus, on the whole, implicit maze training within the comprehension domain can generalize to other types of tasks when the content is similar. These results also suggest that more complex maze training cannot only build the explicit knowledge base, but also aids in developing production skills.

In regards to the grammaticality judgment task post-test, Experiment 2 and 3 demonstrated that it was the Spa training group that showed significantly greater gains. Experiment 2 showed better performance on only ES sentence types, which suggested that the Spa training group may have acquired the same knowledge as the Eng training group, but the difference is that they also displayed better *performance*. This could imply that training with more complex sentences can actually help with performance on easier constructions when the testing measure is explicit. This notion holds many implications for educational programs that focus on developing the four basic skills in the foreign language (listening, speaking, reading, and writing). Experiment 3 showed that the Spa training group yielded fewer errors on their sentence types (SS) as compared to Eng training, which highlights the benefit of the more complex training. Having results that illustrate how the maze task can generalize to another activity that is not implicit itself lends support for its potential as a pedagogical instrument.

Limitations

Some limitations do exist in the present study, and are important to note. Firstly, only an elementary level of Spanish learners was used for participants in the experiments. Because material was appropriate for their level, this is not a surprise. Nevertheless, it would be important to replicate these findings with more advanced learners with level-appropriate sentence constructions. Secondly, all incorrect alternatives were obviously ungrammatical and a different part of speech. With more advanced learners, it would be interesting to see if the maze task can train very subtle grammatical rules (i.e. the indicative vs. the subjunctive). Thirdly, the maze task is completely visual and therefore it would be interesting to see if audio could also be used to enhance learning. For example, with the use of headphones, if the two alternatives are also spoken in one's ears, it would be worthwhile to investigate through a pre-test/post-test design whether listening skills have improved. By reflecting on these limitations, future research with the maze

task can evolve to incorporate a more in-depth investigation with respect to the extent of maze training effects.

CONCLUSION AND FURTHER RESEARCH

The maze task has the potential of fitting into a pedagogical framework because it yields benefits for implicit and explicit knowledge bases alike, which are both instrumental for L2 acquisition to take place. Referring back to Van Patten (2004) and his work with input processing, the maze task can act as implicit practice that is used in conjunction with classroom instruction. The input from the maze can act as a method to draw attention to grammar and lexical meaning assignment, word by word, and in this way could help change L1 preferences if needed. Moreover, the type of training utilized (Spa or Eng) is important with respect to constructing a more solid training paradigm.

This paper has provided further support for the hypothesis that it may be possible to utilize implicit training in order to yield benefits on explicit measures. Moreover, comprehension-based training with this task can generalize to production skills. These findings are of utmost importance when constructing pedagogy that focuses on communication as the primary goal of instruction. Task-based activities such as the maze task, which work to elicit a target L2 structure as its end product, are cornerstones of the communicative foreign language classroom (Richards, 2001). Thus, in order to better place the maze task within a pedagogical framework, a maze utilizing stories is being developed, which will incorporate contextualized sentences. That is, the present study's methodology will be applied in investigating training effects when groups of sentences comprise a story (rather than only using disjointed sentences as was done in this study). In this way, a more complete computer assisted language learning program will be developed and can be used to assist successful foreign language learning.

REFERENCES

Baayen, R.H. (2008a). *Analyzing linguistic data: A practical introduction to statistics*. Cambridge, UK: Cambridge University Press.

Baayen, R.H. (2008b). *languageR: Data sets and functions with "Analyzing Linguistic Data: A practical introduction to statistics"*. R package version 0.953.

Baayen, R.H., Davidson, D.J., & Bates, D.M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, *59*, 390–412.

DeKeyser, R. (1997) Beyond explicit rule learning: Automatizing second language morphosyntax. *Studies in Second Language Acquisition, 19*, 195-221.

DeKeyser, R. & Juffs, A. (2005). Cognitive considerations in L2 learning. In E. Hinkel (Ed.) *Handbook of research in second language teaching and learning* (pp. 437-454). Mahwah, New Jersey: Lawrence Erlbaum Associates.

Ellis, N.C. (1994). Implicit and explicit language learning: An overview. In N. C. Ellis (Ed.) *Implicit and explicit learning of languages* (pp. 1-32). San Diego, CA: Academic Press.

Ellis, N.C. (2005). At the interface: Dynamic interactions of explicit and implicit language knowledge. *Studies in Second Language Acquisition, 27*, 305-352.

Ellis, R. (1993). A theory of instructed second language acquisition. In N. C. Ellis (Ed.) *Implicit and explicit language learning* (pp. 79-114). San Diego, CA: Academic Press.

Ellis, R. (2002). Does form-focused instruction affect the acquisition of implicit knowledge? A review of the research. *Studies in Second Language Acquisition, 24*, 223-236.

Ellis, R. (2009a). Implicit and explicit learning, knowledge and instruction. In R. Ellis, S. Loewen, C. Elder, R. Erlam, J. Philp, & H. Reinders (Eds.) *Implicit and explicit knowledge in second language learning, testing, and teaching* (pp. 3-25). Tonawanda, New York: Multilingual Matters.

Ellis, R. (2009b). Measuring implicit and explicit knowledge of a second language. In R. Ellis, S. Loewen, C. Elder, R. Erlam, J. Philp, & H. Reinders (Eds.) *Implicit and explicit knowledge in second language learning, testing, and teaching* (pp. 237-261). Tonawanda, New York: Multilingual Matters.

Forster, K.I. & Forster, J.C. (2003). DMDX: A windows display program with millisecond accuracy. *Behavior Research Methods, Instruments, & Computers, 35* (1), 116-124.

Forster, K.I., Guerrera, C., & Elliot, L. (2009). The maze task: measuring forced incremental sentence processing time. *Behavior Research Methods, 41* (1), 163-171.

Gasparini, S. (2004). Implicit versus explicit learning: Some implications for L2 teaching. *European Journal of Psychology of Education, 19* (2), 203-219.

Hulstijn, J.H. & de Graaff, R. (1994). Under what conditions does explicit knowledge of a second language facilitate the acquisition of implicit knowledge? A research proposal. *AILA Review*, *11,* 97-112.

Krashen, S. (1994). The input hypothesis and its rivals. In N. C. Ellis (Ed.), *Implicit and explicit language learning* (pp. 45-78). San Diego, CA: Academic Press.

Loewen, S. (2009). Grammaticality judgment tests and the measurement of implicit and explicit L2 knowledge. In R. Ellis, S. Loewen, C. Elder, R. Erlam, J. Philp, & H. Reinders (Eds.) *Implicit and explicit knowledge in second language learning, testing, and teaching* (pp. 94-112). Tonawanda, New York: Multilingual Matters.

Morgan-Short, K. & Bowden, H.W. (2006). Processing instruction and meaningful output-based instruction: Effects on second language development. *Studies in Second Language Acquisition*, *28*, 31-65.

Nitschke, S., Kidd, E., & Serratrice, L. (2009). First language transfer and long-term structural priming in comprehension. *Language and Cognitive Processes, 25* (1), 94-114.

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

Pinheiro, J.C., & Bates, D.M. (2000). *Mixed-effects models in S and S-PLUS*. New York: Springer.

R Development Core Team. (2009). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria.

Richards, J.C. (2001). Course planning and syllabus design. In J. C. Richards (Ed.) *Curriculum development in language teaching* (pp. 145-197). New York, New York: Cambridge University Press.

Thompson, C.K., Shapiro, L. P., & Sobecks, S.K.J. (2003). *Journal of Speech, Language, and Hearing Research, 46*, 591-607.

Tokowicz, N. & MacWhinney, B. (2005). Implicit and explicit measures of sensitivity to violations in second language grammar: An event-related potential investigation. *Studies in Second Language Acquisition, 27*, 173-204.

Van Patten, B. (2004). Input processing in SLA. In B. Van Patten (Ed.) *Processing instruction: Theory, research, and commentary* (pp. 5-32). Mahwah, New Jersey: Lawrence Erlbaum.

Van Patten, B. & Cadierno, T. (1993). Input processing and second language acquisition: A role of instruction. *The Modern Language Journal*, *77* (1), 45-57.

Vygotksy, L.S. (1978). Interaction between learning and development. In M. Cole, V. John-Steiner, S. Scribner, & E. Souberman (Eds.), *Mind in*

society: The development of higher psychological processes (pp. 52-57). Cambridge, MA: Harvard University Press.

Wong, W. (2004). The nature of processing instruction. In B. Van Patten (Ed.) *Processing instruction: Theory, research, and commentary* (pp. 33-64). Mahway, New Jersey: Lawrence Erlbaum Associates.

APPENDIX

Maze Task Questionnaire

Below is a questionnaire in which you can give me your opinion on the maze task!
Please rate your experience from 1 to 5, with 5 being the best rating, and 1 being the worst rating.

5= strongly agree, 4=agree, 3=neutral, 2=disagree, 1=strongly disagree

1) Did you find this task to be fun?

YES NO
5 4 3 2 1

2) Did you find this task to be helpful?

3) YES NO
5 4 3 2 1

4) Do you find this task more enjoyable than online workbook assignments?

YES NO
5 4 3 2 1

5) Did you find this task helpful for your Spanish learning?

YES NO
5 4 3 2 1

6) Did you find yourself *wanting* to try the sentence again if you got it wrong rather than just passing through it?

YES NO
5 4 3 2 1

7) Do you think that this task can help others learn Spanish?

YES NO
5 4 3 2 1

8) Do you think that this task could be helpful for other languages?

YES NO
5 4 3 2 1

9) Do you think this type of practice carries over to doing better on exams/papers, etc.?

YES NO
5 4 3 2 1

10) Did you think that the first session was just as fun as the fourth? (In other words, did it get old fast or do you think you could really get into it for a whole semester?)

YES, I can get into it NO, it got boring
5 4 3 2 1

11) Do you think that if there was a tally of reaction times (that is, how fast you are going), would this increase the fun factor of getting the answer correct?

YES NO
5 4 3 2 1

12) Do you think that this would be a good addition to the Spanish curriculum?

YES NO
5 4 3 2 1