VIAL

Vigo International Journal
of Applied Linguistics

Universidade Vigo

VIAL. Vigo International Journal of Applied Linguistics.

VIAL is an international and peer-reviewed journal which is indexed and abstracted in CINDOC (Centro de Información y Documentación Científica), EBSCO, ERIH PLUS, JCR, Web of Knowledge (Arts and Humanities Citation Index, Social Sciences Citation Index), Linguistics Abstracts, Linguistics and Language Behavior Abstracts, Latindex, MLA Directory of Periodicals, PsychoInfo, Scimago Journal Rank Scopus, Ulrich's Database

Webpage: www.webs.uvigo.es/vialjournal/

Printed in Spain - Impreso en España
I.S.S.N.: 1697-0381 - Depósito Legal: VG-935-2003
Imprime e maqueta: Tórculo Comunicación Gráfica, S.A. Deseño da portada: Elena Gómez.

VIAL

Vigo International Journal of Applied Linguistics

Number 17 - 2020

Contents:

Attention-direction versus retrieval practice: which fosters the productive recall of German formulaic sequences best?

Griet Boone
Griet.Boone@UGent.be

June Eyckmans
June.Eyckmans@UGent.be
Department of Translation, Interpreting and Comunication
Universiteit Gent, Belgium

Abstract

This article reports on a classroom-based (quasi)-experiment with a pre-test post-test design that explored the effect of two types of activities on the productive recall of German formulaic sequences (FS): 1) attention-directing activities and 2) retrieval practice. Two intact classes of Dutch-speaking university students of German participated in the study. One class was randomly assigned to the attention-directing condition (n=18), the other one to the retrieval condition (n=11). Twenty-two target FS were selected as learning items. Each group processed the FS in a different condition. In the attention-directing condition, students had to 1) re-read a video transcript with the FS in bold typeface and 2) translate the targets into Dutch. In the retrieval condition, students had to 1) complete a transcript in which the FS were deleted and 2) translate the targets into German. Results indicate that the retrieval condition led to better productive phrase learning than the attention-directing condition.

Keywords: formulaic sequences; foreign language acquisition; language teaching; attention-directing; retrieval practice

Zusammenfassung

In diesem Artikel wird ein (Quasi)-Experiment im Fremdsprachenunterricht mit einem Prätest-Posttest-Design vorgestellt, das die Auswirkungen zweier Arten von Aktivitäten auf die produktive Wiedergabe deutscher formelhafter Sequenzen (FS) untersuchte: 1) Aufmerksamkeitsfokussierungsaktivitäten und 2) aktivem Gedächtnisabruf (engl. *retrieval practice*). Zwei Lerngruppen von niederländischsprachigen DaF-Studierenden nahmen an der Studie teil. Achtzehn Studierenden wurde

willkürlich die Lerngruppe mit Aufmerksamkeitsfokussierungsaktivitäten zugewiesen, elf Studierenden die Lerngruppe mit Gedächtnisabrufaktivitäten. Zweiundzwanzig formelhafte Sequenzen wurden als Zielobjekte ausgewählt und jede Lerngruppe behandelte die FS in einer anderen Lernumgebung. In der Lernumgebung der Aufmerksamkeitsfokussierung mussten die Studierenden 1) das Transkript mit fettgedruckten FS lesen und 2) die Zielobjekte ins Niederländische übersetzen. In der Lernumgebung mit den Gedächtnisabrufaktivitäten mussten sie 1) das Transkript, in dem die FS vorher gelöscht wurden, als Lückentext ergänzen und 2) die Zielobjekte ins Deutsche übersetzen. Die Ergebnisse deuten darauf hin, dass aktiver Gedächtnisabruf zu einem produktiveren Lernen formelhafter Sequenzen führt im Vergleich zu den Aufmerksamkeitsfokussierungsaktivitäten.

Stichwörter: formelhafte Sequenzen; Fremdsprachenerwerb; Sprachdidaktik; Aufmerksamkeitsfokussierung; aktiver Gedächtnisabruf

1. Introduction

In the last decennia, formulaic sequences (FS) have been a popular object of study, not only in first and second language (L1 and L2) acquisition research, but also in grammatical theory, psycholinguistics and corpus linguistics (Wray, 2009). This has led to a plethora of terms for this phenomenon: *multiword units, lexical chunks, formulas, prefabs, lexical phrases* are but a few of the terms that have been put forward in the literature.

Mastery of formulaic sequences, in all their various guises (i.e. collocations, idioms, proverbs and so forth) has been shown to be an essential component of successful language learning and use (e.g. Meunier & Granger, 2008; Sinclair, 1995; Wray, 2002). Apart from the general consensus about the importance of FS in foreign language learning and teaching, it is also widely accepted that they are a stumbling block for L2 learners (e.g. Conklin & Schmitt, 2008; Laufer & Waldman, 2011).

Different studies have demonstrated that L2 learners need help when learning FS and that the development of a repertoire of FS needs to be supported by language instruction processes (e.g. Meunier, 2012; Szudarski, 2017). Several pedagogical attempts to promote the teaching of L2 phrases have been put forward since the 1990s. However, there is still a need for more empirical research on the effectiveness of pedagogical techniques with a view to language production. Furthermore, the majority of the studies in this field focus on English as a second or foreign language, on figurative idioms or on academic FS. Studies on teaching formulaic language in other languages such as German, the most widely spoken mother tongue in the

European Union, or studies in which non-academic FS from natural discourse are integrated in classroom activities, are rare. The current study aims to explore the effect of two pedagogical activities, namely attention-directing or awareness-raising activities and retrieval practice, on the productive recall of German FS in a classroom context. Attention-direction is a well-established approach for teaching FS in the foreign language classroom, but most studies focus on its effect on recognition, not on production. Retrieval practice is a technique that has been shown to foster productive knowledge of L1 and L2 vocabulary, but that has hardly been the focus of research when it comes to FS. The aim of the present study is, therefore, to weigh the efficiency of attention-directing activities and retrieval practice for L2 phrase learning against each other.

2. Literature review

2.1. Formulaic sequences in foreign language acquisition

One of the most cited definitions of a formulaic sequence is undoubtedly the one by Wray (2002:9): "a sequence, continuous or discontinuous, of words or other elements, which is, or appears to be, prefabricated: that is, stored and retrieved whole from memory at the time of use, rather than being subject to generation or analysis by the language grammar." Wray (2008) makes a clear distinction between a speaker-external approach to formulaicity (i.e. what is formulaic in the language, for example collocations like *raise doubts*, idiomatic expressions like *once in a blue moon*, pragmatic formulas like *have a nice day*, etc.) and a speaker-internal or psycholinguistic approach (i.e. what is formulaic for an individual learner, namely which sequences are stored holistically and retrieved more easily than others by this learner). Especially in studies with L2 learners, where the focus mainly lies on the appropriate use of collocations, idioms, etc., a speaker-external approach to formulaicity is often adopted (e.g. Laufer & Waldman, 2011; Nesselhauf, 2003; Paquot & Granger, 2012). This also applies to this study, in which we define FS as *linguistic clusters* (Myles & Cordier, 2017:10): "multimorphemic clusters which are either semantically or syntactically irregular, or whose frequent co-occurrence gives them a privileged status in a given language as a conventional way of expressing something."

In foreign language acquisition, mastery of formulaic sequences has been shown to be pivotal for successful language learning and use (e.g. Ellis & Simpson-Vlach, 2009; Meunier & Granger, 2008; Nattinger & DeCarrico, 1992; Schmitt, 2004; Sinclair, 1995; Wray, 2002). First, research in corpus linguistics has revealed that FS are widespread in spoken and written native discourse (e.g. Erman & Warren,

2000; Meunier, 2012; Sinclair, 1995). FS are therefore essential if one aims to reach a native-like level of proficiency, especially with a view to the production of idiomatic language (Cowie, 1992; Pawley & Syder, 1983). Using FS has been shown to reduce the cognitive load that L2 processing poses on learners, leading to a more fluent and accurate L2 speech (e.g. Ellis, Simpson-Vlach, & Maynard, 2008; Wood, 2006, 2010, 2012). Finally, there is considerable evidence that L2 learners' language production is considered as more proficient when FS are used, in both L2 speaking and writing (e.g. Boers, Eyckmans, Kappel, Stengers, & Demecheleer, 2006; Ellis, 2001; Nesselhauf, 2003).

Apart from the studies that demonstrate the importance of FS for foreign language learning, there is general consensus that the acquisition of formulaic language is slow (e.g. Boers, Lindstromberg, & Eyckmans, 2014; Laufer & Waldman, 2011) and difficult for language learners (e.g. Bahns & Eldaw, 1993; Conklin & Schmitt, 2008), especially at a productive level (Peters, 2016). It has been shown that L2 language use is characterized by an overuse, underuse, and/or misuse of native-like chunks, even at advanced levels (Nesselhauf, 2003; Durrant & Schmitt, 2010). Many reasons have been put forward to explain the fact that FS are a stumbling block. First, a correct use of FS requires a sensitivity to native speakers' preferred word combinations, and learners usually do not have sufficient exposure to be able to recognize and process these units as recurring lexical units (e.g. Durrant & Schmitt, 2010; Wray, 2000; Yamashita & Jiang, 2010). Second, FS are often not noticed, because learners do not tend to focus their attention on the sequence as a whole but rather on individual words (Barfield & Gyllstad, 2009). Moreover, language learners often assume that FS in other languages are similar to those in their mother tongue and consequently L2 learners produce them through a process of L1 transfer, resulting in erroneous FS in the target language (e.g. Biskup, 1992; Granger, 1998; Nesselhauf, 2003; Paquot & Granger, 2012).

2.2. Teaching formulaic sequences

From the previous, it is clear that L2 learners need support in acquiring FS. Fortunately, several pedagogical attempts to promote the teaching of L2 phrases have been put forward. In the nineties, Willis (1990), Nattinger and DeCarrico (1992) and Lewis (1993) made recommendations on how to introduce FS in the classroom, although only limited empirical research on how to teach FS effectively was available. Twenty years later, Boers and Lindstromberg (2012) presented an extensive review of experimental and intervention studies on pedagogical treatments that are likely to be helpful for L2 learners' acquisition of FS. One of the techniques that has been shown to be effective, is directing learners' attention to formulaic sequences. These attention-directing activities are in line with Lewis' Lexical Approach (1993)

and are focused at raising learners' awareness of the lexical nature of language. The idea behind it is that learners would start noticing FS outside the classroom, which would then foster autonomous learning of FS. Other scholars have used attention-direction through text chunking (e.g. Boers et al., 2006; Jones & Haywood, 2004; Stengers, Boers, Housen, & Eyckmans, 2010) or typographic enhancement (e.g. Boers, Demecheleer, He, Deconinck, Stengers, & Eyckmans, 2017; Choi, 2017; Peters, 2012; Sonbul & Schmitt, 2013; Szudarski & Carter, 2016) and have revealed beneficial effects. Drawing learners' attention to prevalent sound repetition in FS (alliteration, assonance, rhyme) has also been shown effective for recalling formulaic sequences (Boers & Lindstromberg, 2005; Boers, Lindstromberg, & Eyckmans, 2012; Eyckmans, & Lindstromberg, 2017; Lindstromberg, & Eyckmans, 2014), as well as using mental imagery to help learners remember the meaning of figurative idioms (Steinel, Hulstijn & Steinel, 2007; Szczepaniak & Lew, 2011).

Recent studies have demonstrated that explicit teaching of FS results in learning gains. Dictogloss exercises for example have shown to be promising to promote the learning of academic FS and collocations (e.g. Lindstromberg, Eyckmans, & Connabeer, 2016; Snoder & Reynolds, 2019). Pérez Serrano (2018) investigated which type of instruction fosters chunk recognition and compared attention-directing techniques with explicit vocabulary exercises. In both approaches, participants showed learning gains, but the ones obtained through explicit exercises were higher than those obtained through attention-directing techniques. Webb and Kagimoto (2009) compared the effects of receptive and productive vocabulary tasks on learning collocations and showed that both tasks led to significant learning gains. Peters and Pauwels (2015) investigated the effect of different activities on students' recognition, cued output and spontaneous use of academic FS. Students were offered different types of activities: recognition activities, cued output activities or a combination of both. The authors' tentative conclusion is that activities containing cued output might be more beneficial than recognition activities, both at a productive and receptive level. In a study of Laufer and Girsai (2008), in which they examined the effect of content-oriented tasks, text-based vocabulary tasks and translation tasks (L1-L2 and L2-L1) on the acquisition of collocations, translation exercises were found to be most effective.

Overall, many researchers agree that FS are fundamental to language learning and use, and, as a consequence, need to be explicitly addressed in teaching. As there is still a need for more empirical research on the effectiveness of pedagogical techniques, the present study aims to explore the effect of two pedagogical activities, namely attention-directing techniques and retrieval practice, on the productive recall of German FS.

Drawing the attention to the form of the L2 is known to be necessary for successful language learning (e.g. Robinson, 1995; Schmidt, 1990). As mentioned before, there

is a growing number of studies that employ awareness-raising activities with a view to FS learning. However, the majority of the studies focus on receptive recall, and only a few on productive recall. One study that focused on (oral) production of FS and that was carried out during a general proficiency course of 22 teaching hours, showed that students who were made aware of L2 word combinations and engaged in frequent chunking activities in the classroom, used significantly more FS in their conversations in comparison with a control group, engaged in more traditional (grammar-vocabulary) classroom activities (Boers et al., 2006). Another study that focused on the productive recall of FS is a study by Peters (2012), in which two attention-drawing techniques were compared: 1) directing the attention of the learners to FS in a text (instructional method) and 2) typographic salience (bold typeface and underlined). It was demonstrated that typographic salience of the targets had a positive effect on L2 learners' recall of the FS, whereas directing learners' attention did not.

Retrieval practice, also known as the "testing effect", involves the active recall of information from memory. This means that after a learning phase, learners engage in a series of activities (e.g. take a quiz or a test), where they are required to retrieve the learned knowledge, which is supposed to demand a lot of effort from the learner (Karpicke, 2017). As Karpicke (2017:5) writes: "Effortful retrieval of knowledge leaves that knowledge strengthened, increasing the likelihood that it can be accessed and used again in the future." Retrieval practice has proven to be a powerful tool for long-term retention (e.g. Karpicke & Roediger, 2008; Roediger & Butler, 2011) and enhances learning more than does repeated rehearsal, for example (Roediger & Karpicke, 2006). Benefits have been demonstrated in different areas of study and among researchers there is a strong consensus that retrieval practice is beneficial to learning foreign vocabulary (e.g. Barcroft, 2007; Goossens, Camp, Verkoeijen, & Tabbers, 2014; Kang, Lindsey, Mozer, & Pashler, 2014; Karpicke & Smith, 2012; van den Broek, Takashima, Segers, & Verhoeven, 2018). However, these studies concern individual word learning with word lists or word pairs, and not the uptake of L2 FS. To our knowledge, only the study of Peters and Pauwels (2015) mentions retrieval explicitly as a learning activity to acquire FS. They state that in their study, the cued output activities (gap filling and rephrasing activities) can be considered as "retrieval" activities.

3. The study

3.1. Research questions and hypotheses

As we have seen in the literature review, there is some evidence that language students learn FS effectively through attention-drawing techniques, although benefits

were attested mainly with reference to receptive knowledge. The present study will focus on productive knowledge. The effect of attention-directing activities (= typographic salience and translation into L1) on productive recall will be explored. The combination of both activities is intended to raise the chance that the FS are noticed and retained. The attention-directing condition will be compared to a retrieval condition, because this is a method that calls for further investigation when it comes to the learning of FS. The aim of the study was twofold: to see which technique is more powerful to learn lexical phrases at a productive level and to investigate which technique is more beneficial for long-term retention.

The following research questions were addressed:

(1) Which of these conditions (attention-direction or retrieval practice) leads to the largest uptake of FS?

(2) Is the difference in uptake maintained over time?

Because the findings regarding the benefits for single word acquisition through retrieval practice are quite robust, we hypothesize that retrieval practice will lead to higher learning gains than the attention-directing activities. Along those same lines, we predict that learners of the retrieval condition group will be able to produce more target FS in a delayed post-test, two months after the intervention.

3.2. Methodology

3.2.1. Design

A classroom-based controlled (quasi)-experiment was set up and two intact classes were randomly assigned to either an attention-directing condition or a retrieval condition. A combination of a within-subject and a between-subject design with a pre-test, immediate post-test and two delayed post-tests was adopted. To guarantee that results would be directly comparable, the target items tested in the pre-, post-, and delayed post-tests were identical. Students were not informed of the fact that they would be tested on their productive knowledge of the German FS.

3.2.2. Participants

Participants in this study were two intact classes of Dutch-speaking students in their second bachelor year of an Applied Linguistics program at a large Belgian

university, who were majoring in German and an additional foreign language of their choice. Their ages ranged between 19 and 22. The experiment was conducted during a German course that aims to improve students' oral production in German. This course consisted of a weekly two-hour class over a period of 12 weeks. All students received 190 contact hours of formal instruction in German before the start of the study and their proficiency level for German was assessed at the B1 level for production and at the B2 level for comprehension according to the Common European Framework of Languages (Council of Europe, 2001). The students were administered the Productive Vocabulary Test for German, developed by the Institute for Test Research and Test Development, in cooperation with the Herder-Institute Leipzig and the University of Leipzig. The vocabulary levels of German (1000, 2000, 3000, 4000 and 5000) are based on the frequency lists developed from the Herder/BYU-corpus (Jones, Tschirner, Goldhahn, Buchwald, & Ittner, 2006). To pass the test, participants need to score 14 out of 18 items per level. All students, except for one, passed the test for the 1000 level, and almost two-thirds passed the test for the 2000 and 3000 level. Nobody reached the 4000 or 5000 level, which means that the participants constitute a rather homogenous group in terms of their level of productive vocabulary knowledge.

Thirty-four students participated in the session in which the pre-test (i.e. a gap-fill-exercise, including the 22 target items), the pedagogical intervention and the immediate post-test (i.e. a sight translation, including the same 22 target items) were carried out. One week later, 32 students participated in the delayed post-test, which consisted of the same gap-fill-exercise as in the pre-test. One student was excluded from the experiment, because she made a wordlist during class and used it to complete the immediate and delayed post-test. After two months, the same post-test was administered to both groups, without prior notice. Only 29 students (5 male, 24 female) attended class this time, resulting in 18 participants for the attention-directing condition and 11 for the retrieval condition.

3.2.3. Target items

For this study, 22 target items were selected from the transcript of a German video-recording, a text of 335 words. This short informative German video-recording was chosen for three reasons: (1) Authenticity: as it was taken from ARD, the German public broadcaster, it was not specifically designed for teaching purposes. In other words, the FS used in the video-recording were part of authentic German spoken discourse and meet Nesselhauf´s criterion for the selection of FS to be taught, i.e. "acceptable and frequent in a neutral register" (Nesselhauf, 2003:238); (2) Content: the topic of legalizing marijuana lent itself well to a class-room discussion; (3) Length: the

video-recording took less than three minutes to watch and was therefore sufficiently short to keep the students interested.

On the basis of the transcript of the video-recording, 22 items (see Appendix A) were selected as targets according to the following criteria: 1) the sequence contains at least two words; 2) the sequence is either listed in one of the two existing German collocation dictionaries: *Feste Wortverbindungen des Deutschen: Kollokationenwörterbuch für den Alltag* (Häcki Buhofer, Dräger, Meier, & Roth, 2014) and *Wörterbuch der Kollokationen im Deutschen* (Quasthoff, 2011), or in the German newspaper corpus of the Leipzig Corpora Collection (LCC), which can be accessed online at http://corpora.uni-leipzig.de/.

3.2.4. Test instruments

To test students' productive knowledge of the targeted FS, a gap-fill exercise was designed to be used as a pre-test and a delayed post-test. Twenty-two target FS were left out from the full-length transcription text and students were asked to fill in the blanks. For all items, students' L1 (Dutch) translation was given as a prompt. In some cases, the first letter of a (part of the) target item was given to exclude other possible response alternatives.

As an immediate post-test, a sight translation task from Dutch into German was chosen. In this kind of task, students read the text in the L1 and translate it out loud in the L2. The sight translation task solicited the 22 target items and was selected because of the participants' familiarity with this task and its unequivocal productive nature. The gap-fill exercise (pre-test and delayed post-tests) can be found in Appendix B, the immediate post-test in Appendix C.

3.2.5. Procedure

For each condition (attention-direction and retrieval), two types of activities were designed. The attention-directing activities consisted of 1) reading the authentic transcription text with typographic salient target FS and 2) engaging in a flash-card exercise in which the FS were translated into L1 Dutch. The retrieval activities consisted of 1) completing the FS that were left out of the transcript (gap-fill exercise) and 2) engaging in a flash-card exercise in which the FS were translated into L2 German.

The pre-test, the pedagogical intervention and the immediate post-test were performed in the fifth week of the 12-week course. In this week, both parallel groups attended an identical two-hour class, given by the same lecturer. After a short

introduction about the topic and objective of the class, all students completed the pre-test. They got five to ten minutes to fill in the 22 missing targets. The pre-test was announced as an introductory vocabulary exercise to see which FS were already familiar and which lexical items students would learn that day.

After all pre-tests were collected, students viewed the video-recording. Because it is known that repetition in vocabulary-focused activities is beneficial for word learning (Nation, 2013) and that the same applies for FS (Alali & Schmitt, 2012), students were exposed to the 22 target items four times: twice when watching the video-recording and twice in the exercises that followed.

To verify students' comprehension of the video-recording, content questions were asked and answered after the first viewing. Students were encouraged to concentrate on the vocabulary and the FS used when watching the video-recording a second time. All students were already familiar with the concept of FS and the advantages of mastering FS when learning a foreign language.

Immediately after watching the video-recording, both groups received a first exercise targeting the 22 FS. In the attention-directing condition group, the transcript of the video-recording was projected on a screen and participants read the text. The 22 FS were in bold typeface and the teacher directed the attention of the learners to these FS, asking students to translate them into Dutch. In the retrieval condition group, the transcript was projected as a gap-fill exercise, in which the 22 FS were left out. Participants had to read the text and fill in the blanks orally. Both groups received corrective feedback if a wrong (or no) answer was provided.

During the remainder of the class, students were invited to exercise their speaking skills through the discussion of other short texts. Approximately 20 minutes before the end of class, there was a second exercise to learn the 22 FS. Both groups engaged in a flashcard exercise: for all 22 targets, digital flashcards were created, using *Quizlet*. On one side of the flashcard, the FS was in Dutch, on the other side in German. Students in the attention-directing condition saw the phrases in German and translated them orally into Dutch; students in the retrieval condition translated the 22 FS from L1 Dutch into L2 German.

At the end of class, students of both groups completed the same immediate post-test, a sight translation from Dutch into German, containing the 22 FS. They made recordings of their oral translations and these were uploaded on the learning platform of the course. The students were accustomed to this procedure.

In the first delayed post-test, one week later (week six of the course), and in the second delayed post-test, two months later (week 12), students of both groups were asked to fill in the blanks of the same gap-fill-exercise they had completed as a pre-test. It took them five to ten minutes.

3.2.6. Scoring and analysis

The pre-test and the two delayed post-tests were scored dichotomously. Partial knowledge (e.g. one correct word of a two-word phrase) was not taken into account, but accurate spelling was a prerequisite.

For the sight translation, the audio recordings of the students were analysed with a focus on the translation of the target phrases. Again, dichotomous scoring was applied: one point for a correct rendering of the phrase in German and zero for an unacceptable phrase. Pronunciation had to be satisfactory to obtain credits. In German, a mistake in the use of an *umlaut* (i.e. vowel alternation) for example, can change the meaning of a word, e.g. *fordern* (to demand) versus *fördern* (to support). For all tests, the maximum score was 22, as there was a total number of 22 target items.

A mixed model repeated measures Analysis of Variance (ANOVA) was conducted. The between-group variable was the type of instruction (attention-direction and retrieval) and the within-group variable was the test score at four different points in time: 1) the pre-test, prior to the pedagogical intervention, 2) the immediate post-test, immediately after the pedagogical intervention, 3) the first delayed post-test, one week after the pedagogical intervention, and 4) the second delayed post-test, two months after the pedagogical intervention. SPSS Statistics 25 was used for the statistical analysis.

4. Results

4.1. Descriptive statistics

Both groups' mean scores and standard deviations on the pre-test and the three post-tests are listed in Table 1. Participants in the attention-directing condition scored slightly higher on the pre-test. There is a treatment effect in both groups: there are noticeable learning gains in the attention-directing condition as well as in the retrieval condition. However, learning gains in the retrieval condition seem to be higher than the learning gains in the attention-directing condition. The descriptive results also show that the attrition over time is slightly smaller in the retrieval practice condition:

on average 2 target phrases were no longer remembered by participants in the retrieval condition, versus 2.39 in the attention-directing condition. What stands out is that students in the retrieval condition remember on average 14 out of 22 items after two months, which corresponds to the score the participants of the attention-directing condition obtained in the immediate post-test.

Table 1. Means (and standard deviations) of the scores on the target phrases for both groups. Maximum score = 22.

	Time			
	pre-test	**immediate post-test**	**delayed post-test 1**	**delayed post-test 2**
Attention-direction (*n*=18)	5.89 (1.90)	14.22 (2.67)	13.11 (3.27)	11.83 (3.01)
Retrieval practice (*n*=11)	5.27 (1.73)	16.45 (2.11)	15.82 (2.82)	14.45 (2.70)

Data were pre-analysed to check the assumptions of normality, homogeneity of variances and sphericity, before applying a mixed ANOVA. The Shapiro-Wilk test showed a normal distribution of the data. The error variance of the dependent variable is equal across groups, based upon results of Levene's test (pre-test, F (1, 27) = .328, p = .572; post-test 1, F (1, 27) = .943, p = .340; post-test 2, F (1, 27) = 1.321, p = .261); post-test 3, F (1, 27) = .016, p = .901). This means that the assumption of homogeneity of variances for the groups was met. The assumption of sphericity, based upon Mauchly's Test of Sphericity, was not violated, p = .314.

4.2. Effect of attention-directing activities and retrieval practice

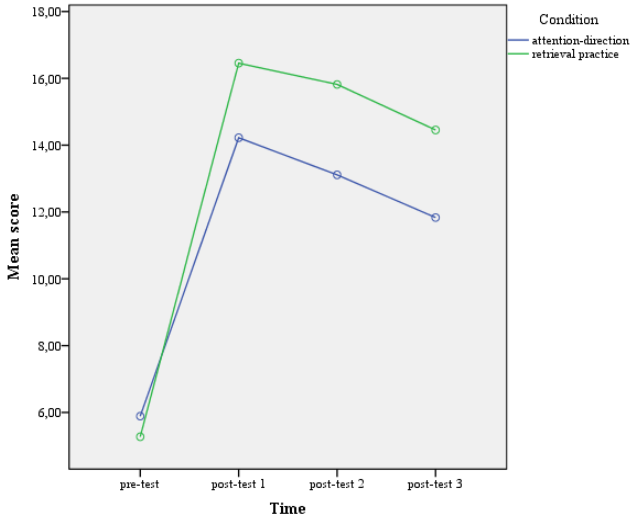
In the mixed ANOVA, the between-subjects factor is condition (attention-direction versus retrieval) and the within-subjects factor is time (with four different time points).

Concerning the between-subjects factor, the data analysis revealed that there was a significant main effect of condition on the learning of target items, F (1,27) = 6.09, p = 0.02, η^2_p = .18. This means that participants' uptake of FS differed according to the condition. Participants of the retrieval condition scored higher on the three post-tests than their peers in the attention-directing condition group.

Concerning the within-subjects analysis, the results demonstrated a statistically significant main effect of time, F (3, 81) = 115.41, $p < 0.001$, $\eta^2_p = .81$. Contrasts revealed that participants' average score on the immediate post-test was significantly higher than on the pre-test, F (1, 27) = 144.63, $p < .001$, $\eta^2_p = .84$ and that their score on the first delayed post-test was again significantly higher than on the pre-test, F (1, 27) = 123.80, $p < .001$, $\eta^2_p = .82$. The score on the second delayed post-test was also significantly higher than on the pre-test, F (1, 27) = 49.58, $p < .001$, $\eta^2_p = .65$.

The analysis points to a statistically significant interaction effect between time and condition, F (3, 81) = 3.63, $p = 0.016$, $\eta^2_p = .12$. This result suggests that the effect of time on the phrasal uptake is different for students engaged in retrieval activities than for students who engaged in attention-directing activities. To break down this interaction, contrasts were performed comparing differences in productive knowledge of the targets across the groups. These revealed a significant interaction when comparing the scores for the pre-test to the scores of the immediate post-test, F (1, 27) = 7.88, $p = .009$, $\eta^2_p = .23$, as is illustrated by the interaction graph (Figure 1). The results of the further contrasts, which compared the immediate post-test and delayed post-tests did not reveal significant interaction effects, F (1, 27) = 2.17, $p = .152$, $\eta^2_p = .07$ and F (1, 27) = 0.40, $p = .536$, $\eta^2_p = .01$. The statistical significance of the first contrast appears to result from the difference between the attention-directing and the retrieval condition, in which the slope of the line between the pre-test and immediate post-test for the retrieval condition group is steeper than for the attention-directing condition group, showing that the learning gains of participants in the retrieval condition are considerably higher immediately after the pedagogical intervention. If we look at the lines of the immediate post-test and the first delayed post-test, we see that this line is flatter in the retrieval condition group, showing that attrition after one week in this group was smaller than in the attention-directing condition group. The attrition after one week is smaller (0.63) in the retrieval condition compared to the attention-directing condition (1.11). Looking at the line between the first and second delayed post-test, there is a similar level of decline in scores between the two groups. During this period between first and second delayed post-test, the attrition in both groups is similar: 1.28 in the attention-directing condition versus 1.37 in the retrieval condition. However, at the final test, two months after the intervention, the mean performance of the retrieval condition (14.45) is at the same level as the mean score of the attention-directing condition (14.22) immediately after the pedagogical intervention.

Figure 1. Test scores of both groups over time



5. Discussion

In the current study, the merits of attention-directing activities and retrieval practice for the productive recall of German FS were compared. The main aim of this study was to see if there is a significant difference in the uptake of FS when learning FS through attention-directing activities or retrieval practice (RQ1), and if this difference in uptake is maintained over time (RQ2). Although the results demonstrate considerable learning gains in both conditions, retrieval practice led to a larger uptake of FS than attention-direction, thereby answering the first research question. Regarding the second research question, the data show that the difference in uptake is maintained over time, with better recall rates for the retrieval practice condition after two months.

With regard to the attention-directing activities, students who engaged in this condition were able to increase their productive knowledge of the 22 targets from 5.89 on the pre-test to 14.22 on the immediate post-test, 13.11 on the delayed post-test and 11.83 on the final post-test. The largest gain (51.7%) is situated between the pre-test and the immediate post-test, which is to be expected. When we look at the attrition of this new knowledge in both delayed tests, the scores the students obtained still show a gain of 44.8% and 36.9% respectively with reference to the pre-test. These results support previous findings on awareness-raising activities when teaching FS: in order to acquire FS, foreign language learners need to notice them. Because students have difficulties noticing or identifying FS in the foreign language, pedagogical support is needed (e.g.

Boers & Lindstromberg, 2012). In this study, the first activity in the attention-directing condition was making the targets more salient through bold typeface. Although students were not aware of the fact that they would be tested afterwards, they still seemed to remember the targets. This is in line with the study of Peters (2012), in which typographic salience was shown to be effective for the recall of FS. As a second activity in the attention-directing condition, participants had to translate the targets from L2 into L1. This activity served to reinforce students' awareness of the form and meaning of the FS. A combination of both attention-directing activities resulted in a recall rate of the targets of 11.83 (out of 22) after two months.

The students in the retrieval practice condition scored 5.27 on the pre-test, 16.45 on the immediate post-test, 15.82 on the delayed post-test and 14.45 on the final post-test. Here too, the largest knowledge gain (66.8%) is attested immediately after the intervention in class, after which attrition sets in. After one week, students still show a gain of 63% with reference to the pre-test and after two months of 54.9%. These results provide further support for the findings of Peters and Pauwels (2015). In their study, the cued output or retrieval activities (gap filling and rephrasing activities) had a positive effect on students' awareness, cued output and spontaneous use of academic FS. Additionally, our results corroborate the findings of available studies on retrieval practice in L2 vocabulary learning (e.g. Barcroft, 2007; Goossens et al., 2014; Karpicke, 2017; Karpicke & Roediger, 2008; van den Broek et al., 2018). More importantly, they provide much-needed evidence of the benefits of retrieval practice for L2 phrase learning.

The superior scores obtained through retrieval practice can be explained by the effort that is required from the students in retrieving and producing the target FS. According to theoretical accounts of retrieval practice, effortful retrieval is needed to acquire durable knowledge (Karpicke, 2017). Completing deleted targets in a text (retrieval condition) is more challenging for students than reading them (attention-directing condition). However, also the attention-directing activity with the highlighted FS, in combination with the focus on the meaning (through the translation into L1) seems to have had a positive effect. This is in line with Schmidt's (1990) noticing hypothesis, which implies that awareness of the form of input at the level of "noticing" is indispensable for foreign language learning. This, in combination with a contrastive engagement with the targets, in which students are made aware of interlingual differences as proposed in the study of Laufer and Girsai (2008), seems to be effective.

The attrition rates were similar across groups. However, the most considerable recall gain was attested in the retrieval practice group. This is consistent with the hypothesis of "pushed output" (Swain, 1985), which claims that learners improve their language development when they produce language and are "pushed" to do so.

Participants in the retrieval practice group were pushed to produce L2-output through both retrieval activities (gap filling and translating from L1 into L2). This seems to have led to successful recall of FS in all post-tests and especially in the immediate post-test.

The attention-directing activities have not required a similar amount of effort from the students. Although students were made aware of the form (by reading the targets in bold typeface) and the meaning of the FS (by translating them from L2 into L1), this has not led to a comparable success in recall of the target FS.

6. Conclusion

In L2 acquisition, mastery of formulaic sequences on a productive level is one of the challenges L2 learners have to cope with. Due to the importance of FS in foreign language acquisition, there is no doubt that FS have to be addressed in class and that teachers should include explicit focus on form when teaching FS.

The comparison of the effect of attention-directing activities and retrieval practice on the acquisition and recall of FS by L2 learners, revealed that students made learning gains in both conditions. Retrieval practice, however, benefited learning more than attention-directing activities, as was evidenced by the significant higher learning gains in the retrieval condition.

There are a number of reasons why these findings should be interpreted with caution. First, due to the nature of our sample with two intact classes and because of the limited number of students learning German as a foreign language, no control condition could be added to the design. Second, the relatively modest sample of participants and the unbalanced size of the groups constrains the external validity of this study. A third limitation concerns the number of selected phrases: 22 target items is a rather limited number, considering the amount of vocabulary that university students, majoring in foreign languages, have to study. However, the range of test scores (between 11 and 18 out of 22 in the retrieval practice group, and between 6 and 17 in the attention-directing condition group) does not point to a ceiling effect.

In future research, individual difference variables such as language aptitude or working memory could be taken into account in order to see which effect different pedagogical conditions have on different learners. Individuals differ in memory and attentional capacity, which influences the extent of noticing and thus has a direct effect on SLA (Robinson, 1995).

7. Pedagogical implications

Based on our findings, we make a plea for a pedagogical practice in which attention-directing activities are taken as a starting point in order to make students aware of the syntagmatic character of the language and to help them identify specific FS in the foreign language. When the aim of the language course is foreign language production, retrieval activities should also be implemented.

With this study we have illustrated that authentic material, such as short video-recordings in the L2, can be quite easily transferred into teaching material without taking up too much preparation time. Teachers can select FS that are useful for learners and turn them into learning targets through gap-fill exercises. Retrieval practice of this sort should be considered a powerful technique in language courses, not only when it comes to the acquisition of individual words, but also when it comes to the production of FS, which is far more challenging for L2 learners.

8. References

Alali, F. A., & Schmitt, N. (2012). Teaching Formulaic Sequences: The Same as or Different from Teaching Single Words? *TESOL Journal*, *3*(2), 153–180.

Bahns, J., & Eldaw, M. (1993). Should we teach EFL students collocations? *System*, *21*(1), 101–114.

Barcroft, J. (2007). Effects of Opportunities for Word Retrieval During Second Language Vocabulary Learning. *Language Learning*, *57*(1), 35–56.

Barfield, A., & Gyllstad, H. (Eds.). (2009). *Researching collocations in another language: Multiple interpretations*. New York: Palgrave Macmillan.

Biskup, D. (1992). L1 Influence on Learners' Renderings of English Collocations: A Polish/German Empirical Study. In P. J. L. Arnaud & H. Béjoint (Eds.), *Vocabulary and Applied Linguistics*, 85–93.

Boers, F., Demecheleer, M., He, L., Deconinck, J., Stengers, H., & Eyckmans, J. (2017). Typographic enhancement of multiword units in second language text: Typographic enhancement of multiword units in second language text. *International Journal of Applied Linguistics*, *27*(2), 448–469.

Boers, F., Eyckmans, J., Kappel, J., Stengers, H., & Demecheleer, M. (2006). Formulaic sequences and perceived oral proficiency: Putting a Lexical Approach to the test. *Language Teaching Research*, *10*(3), 245–261.

Boers, F., & Lindstromberg, S. (2005). Finding ways to make phrase-learning feasible: The mnemonic effect of alliteration. *System*, *33*(2), 225–238.

Boers, F., & Lindstromberg, S. (2012). Experimental and Intervention Studies on Formulaic Sequences in a Second Language. *Annual Review of Applied Linguistics*, *32*, 83–110.

Boers, F., Lindstromberg, S., & Eyckmans, J. (2012). Are Alliterative Word Combinations Comparatively Easy to Remember for Adult Learners? *RELC Journal*, *43*(1), 127–135.

Boers, F., Lindstromberg, S., & Eyckmans, J. (2014). Is alliteration mnemonic without awareness-raising? *Language Awareness*, *23*(4), 291–303.

Choi, S. (2017). Processing and learning of enhanced English collocations: An eye movement study. *Language Teaching Research*, *21*(3), 403–426.

Conklin, K., & Schmitt, N. (2008). Formulaic Sequences: Are They Processed More Quickly than Nonformulaic Language by Native and Nonnative Speakers? *Applied Linguistics*, *29*(1), 72–89.

Council of Europe. (2001).*Common European framework of reference for languages: Learning, teaching, assessment*. Cambridge, U.K: Press Syndicate of the University of Cambridge.

Cowie, A. P. (1992). Multiword Lexical Units and Communicative Language Teaching. In P. J. L. Arnaud & H. Béjoint (Eds.), *Vocabulary and Applied Linguistics*. 1–12.

Durrant, P., & Schmitt, N. (2010). Adult learners' retention of collocations from exposure. *Second Language Research*, *26*(2), 163–188.

Ellis, N. C. (2001). Memory for language. In P. Robinson (Ed.), *Cognition and Second Language Instruction* 33–68.

Ellis, N. C., & Simpson-Vlach, R. (2009). Formulaic language in native speakers: Triangulating psycholinguistics, corpus linguistics, and education. *Corpus Linguistics and Linguistic Theory*, *5*(1).

Ellis, N. C., Simpson-Vlach, R., & Maynard, C. (2008). Formulaic Language in Native and Second Language Speakers: Psycholinguistics, Corpus Linguistics, and TESOL. *TESOL Quarterly*, *42*(3), 375–396.

Erman, B., & Warren, B. (2000). The idiom principle and the open choice principle. *Text - Interdisciplinary Journal for the Study of Discourse*, *20*(1).

Eyckmans, J., & Lindstromberg, S. (2017). The power of sound in L2 idiom learning. *Language Teaching Research*, *21*(3), 341–361.

Goossens, N. A. M. C., Camp, G., Verkoeijen, P. P. J. L., & Tabbers, H. K. (2014). The Effect of Retrieval Practice in Primary School Vocabulary Learning: Effect of retrieval practice. *Applied Cognitive Psychology*, *28*(1), 135–142.

Granger, S. (1998). Prefabricated patterns in advanced EFL writing: Collocations and formulae. In A.P.Cowie (Ed.), *Phraseology: Theory, analysis, and applications* (pp. 145–160). Oxford: Clarendon Press.

Häcki Buhofer, A., Dräger, M., Meier, S., & Roth, T. (2014). *Feste Wortverbindungen des Deutschen: Kollokationenwörterbuch für den Alltag*. Tübingen: Francke.

Jones, M. A., & Haywood, S. (2004). Facilitating the acquisition of formulaic sequences: An exploratory study in an EAP context. In N. Schmitt (Ed.), *Language Learning & Language Teaching*. Vol. 9:269–300.

Jones, R. L., Tschirner, E. P., Goldhahn, A., Buchwald, I., & Ittner, A. (2006). *A frequency dictionary of German: Core vocabulary for learners*. London ; New York: Routledge.

Kang, S. H. K., Lindsey, R. V., Mozer, M. C., & Pashler, H. (2014). Retrieval practice over the long term: Should spacing be expanding or equal-interval? *Psychonomic Bulletin & Review*, *21*(6), 1544–1550.

Karpicke, J. D. (2017). Retrieval-Based Learning: A Decade of Progress. In *Learning and Memory: A Comprehensive Reference*. 487–514.

Karpicke, J. D., & Roediger, H. L. (2008). The Critical Importance of Retrieval for Learning. *Science*, *319*(5865), 966–968.

Karpicke, J. D., & Smith, M. A. (2012). Separate mnemonic effects of retrieval practice and elaborative encoding. *Journal of Memory and Language*, *67*(1), 17–29.

Laufer, B., & Girsai, N. (2008). Form-focused Instruction in Second Language Vocabulary Learning: A Case for Contrastive Analysis and Translation. *Applied Linguistics*, *29*(4), 694–716.

Laufer, B., & Waldman, T. (2011). Verb-Noun Collocations in Second Language Writing: A Corpus Analysis of Learners' English: Verb-Noun Collocations in L2 Writing. *Language Learning*, *61*(2), 647–672.

Lewis, M. (1993). *The lexical approach: The state of ELT and a way forward*. Hove, England: Language Teaching Publications.

Lindstromberg, S., & Eyckmans, J. (2014). How big is the positive effect of assonance on the recall of L2 collocations? *ITL - International Journal of Applied Linguistics*, *165*(1), 19–45.

Lindstromberg, S., Eyckmans, J., & Connabeer, R. (2016). A modified dictogloss for helping learners remember L2 academic English formulaic sequences for use in later writing. *English for Specific Purposes*, *41*, 12–21.

Meunier, F. (2012). Formulaic Language and Language Teaching. *Annual Review of Applied Linguistics*, *32*, 111–129.

Meunier, F., & Granger, S. (Eds.). (2008). *Phraseology in foreign language learning and teaching*. Amsterdam ; Philadelphia: John Benjamins Pub. Co.

Myles, F., & Cordier, C. (2017). Formulaic sequence(FS) cannot be an umbrella term in SLA. *Studies in Second Language Acquisition*, *39*(01), 3–28.

Nation, I. S. P. (2013). *Learning vocabulary in another language*. Cambridge ; New York: Cambridge University Press.

Nattinger, J. R., & DeCarrico, J. S. (1992). *Lexical phrases and language teaching*. Oxford [England] ; New York: Oxford University Press.

Nesselhauf, N. (2003). The Use of Collocations by Advanced Learners of English and Some Implications for Teaching. *Applied Linguistics*, *24*(2), 223–242.

Paquot, M., & Granger, S. (2012). Formulaic Language in Learner Corpora. *Annual Review of Applied Linguistics*, *32*, 130–149.

Pawley, A. & Syder, F.H. (1983). Two puzzles for linguistic theory: Native-like selection and native-like fluency. In Richards J.C. & Schmidt R.W. (Eds.). *Language and communication.* London, New York: Longman. 191-226.

Pérez Serrano, M. (2018). Which type of instruction fosters chunk learning? Preliminary conclusions. *Revista de Lingüística y Lenguas Aplicadas*, *13*(1), 133.

Peters, E. (2012). Learning German formulaic sequences: The effect of two attention-drawing techniques. *The Language Learning Journal*, *40*(1), 65–79.

Peters, E. (2016). The learning burden of collocations: The role of interlexical and intralexical factors. *Language Teaching Research*, *20*(1), 113–138.

Peters, E., & Pauwels, P. (2015). Learning academic formulaic sequences. *Journal of English for Academic Purposes*, *20*, 28–39.

Quasthoff, U. (2011). *Wörterbuch der Kollokationen im Deutschen*. Berlin, New York: De Gruyter.

Robinson, P. (1995). Attention, Memory, and the "Noticing" Hypothesis. *Language Learning*, *45*(2), 283–331.

Roediger, H. L., & Butler, A. C. (2011). The critical role of retrieval practice in long-term retention. *Trends in Cognitive Sciences*, *15*(1), 20–27.

Roediger, H. L., & Karpicke, J. D. (2006). Test-Enhanced Learning: Taking Memory Tests Improves Long-Term Retention. *Psychological Science*, *17*(3), 249–255.

Schmidt, R. W. (1990). The Role of Consciousness in Second Language Learning. *Applied Linguistics*, *11*(2), 129–158.

Schmitt, N. (Ed.). (2004). *Formulaic sequences: Acquisition, processing, and use*. Amsterdam ; Philadelphia: John Benjamins Publ.

Sinclair, J. (1995). *Corpus, concordance, collocation* (3. impr). Oxford: Oxford Univ. Press.

Snoder, P., & Reynolds, B. L. (2019). How dictogloss can facilitate collocation learning in ELT. *ELT Journal*, *73*(1), 41–50.

Sonbul, S., & Schmitt, N. (2013). Explicit and Implicit Lexical Knowledge: Acquisition of Collocations Under Different Input Conditions: Explicit and Implicit Lexical Knowledge. *Language Learning*, *63*(1), 121–159.

Steinel, M. P., Hulstijn, J. H., & Steinel, W. (2007). Second Language Idiom Learning in a paired-associate paradigm: Effects of Direction of Learning, Direction of Testing, Idiom Imageability, and Idiom Transparency. *Studies in Second Language Acquisition*, *29*(03).

Stengers, H., Boers, F., Housen, A., & Eyckmans, J. (2010). Does "chunking" foster chunk-uptake? In S. De Knop, F. Boers, & A. De Rycker (Eds.), *Fostering Language Teaching Efficiency through Cognitive Linguistics*.

Swain, M. (1985). Communicative competence: Some roles of comprehensible input and comprehensible output in its development'. In S. M. Gass & C. G. Madden (Eds.), *Input in Second Language Acquisition* (pp. 235–253). New York: Newbury House.

Szczepaniak, R., & Lew, R. (2011). The Role of Imagery in Dictionaries of Idioms. *Applied Linguistics*, *32*(3), 323–347.

Szudarski, P. (2017). Learning and Teaching L2 Collocations: Insights from Research. *TESL Canada Journal*, *34*(3).

Szudarski, P., & Carter, R. (2016). The role of input flood and input enhancement in EFL learners' acquisition of collocations: L2 input types and acquisition of collocations. *International Journal of Applied Linguistics*, *26*(2), 245–265.

van den Broek, G. S. E., Takashima, A., Segers, E., & Verhoeven, L. (2018). Contextual Richness and Word Learning: Context Enhances Comprehension but Retrieval Enhances Retention: Effects of Contextual Richness on Word Retention. *Language Learning*, *68*(2), 546–585.

Webb, S., & Kagimoto, E. (2009). The Effects of Vocabulary Learning on Collocation and Meaning. *TESOL Quarterly*, *43*(1), 55–77.

Willis, D. (1990). *The lexical syllabus: An new approach to language teaching*. London: Collins ELT.

Wood, D. (2006). Uses and Functions of Formulaic Sequences in Second Language Speech: An Exploration of the Foundations of Fluency. *Canadian Modern Language Review*, *63*(1), 13–33.

Wood, D. (Ed.). (2010). *Perspectives on formulaic language: Acquisition and communication*. London, New York: Continuum.

Wood, D. (2012). *Formulaic language and second language speech fluency: Background, evidence and classroom applications*. London; New York: Continuum.

Wray, A. (2000). Formulaic sequences in second language teaching: Principle and practice. *Applied Linguistics*, *21*(4), 463–489.

Wray, A. (2002). *Formulaic language and the lexicon*. Cambridge, New York: Cambridge University Press.

Wray, A. (2008). *Formulaic language: Pushing the boundaries* (1. published). Oxford: Oxford Univ. Press.

Wray, A. (2009). Future directions in formulaic language research. *Journal of foreign languages*, *32*, 2-17.

Yamashita, J., & Jiang, N. (2010). L1 Influence on the Acquisition of L2 Collocations: Japanese ESL Users and EFL Learners Acquiring English Collocations. *TESOL Quarterly*, *44*(4), 647–668.

Appendix A

German target phrases used in the study with their Dutch and English translation

staatlich kontrolliert	door de overheid gecontroleerd	state controlled
wesentlich höher	aanzienlijk hoger	considerably higher
(bei) jungen Menschen	(bij) jonge mensen	(among) young people
das Risiko erhöhen	de kans verhogen	to increase the risk
das Gehirn schädigen	schade berokkenen aan de hersenen	to damage the brain
die Wahrscheinlichkeit liegt bei	de kans bedraagt	the probability is around
abhängig werden	verslaafd raken	to become dependent
zum Vergleich	ter vergelijking	for comparison
mehr als doppelt so hoch	meer dan dubbel zo hoog	more than twice as high
nach dem Alter fragen	naar de leeftijd vragen	to ask someone´s age
in Kontakt kommen mit	in contact komen met	to come into contact with
harte Drogen	Harddrugs	hard drugs
zum großen Teil	Grotendeels	in large part
enorme Kosten	enorme kosten	huge costs
Kosten einsparen	kosten besparen	to decrease costs
die Rede sein von	sprake zijn van	to speak of
ordentlich verdienen	flink verdienen	to earn decently
gut angelegtes Geld	goed geïnvesteerd geld	money well spent
die Nachfrage senken	de vraag doen dalen	to reduce demand
ein Umdenken fordern	een mentaliteitsverandering eisen	to demand a change of mindset
Schaden und Nutzen	nadelen en voordelen	advantages and disadvantages
Antworten finden auf die Frage	antwoorden vinden op de vraag	to find answers to the question

Appendix B

Gap-fill exercise (pre-test and both delayed post-tests)

Soll Cannabis legalisiert werden?

Schätzungsweise jeder vierte Deutsche hat schon mal gekifft. Und das, obwohl Cannabis illegal ist. Trotz Strafverfolgung, Razzien und Kriminalisierung: der Cannabiskonsum bleibt auf stabilem Niveau. Dass Besitz, Anbau und Handel verboten sind, ist offenbar wirkungslos. Sollte Cannabis also legalisiert werden?

Deutschland wäre nicht das erste Land. Beispiel: der US-Bundesstaat Colorado: hier ist Cannabis komplett legal. Anbau, Verkauf und Konsum: alles (door de overheid gecontroleerd) s_____ _____. Der Anteil der Cannabiskonsumenten an der Bevölkerung ist hier übrigens nicht oder nicht (aanzienlijk hoger) w_____ _____ als in anderen Ländern, in denen Cannabis illegal ist.

Trotzdem: Cannabis ist eine Droge. Ihr Wirkstoff THC kann vor allem bei (jonge mensen) j_____ M_____ (de kans op … verhogen) _____ Psychose-_____ _____ und (schade berokkenen aan de hersenen) _____. Und, (de kans) d_____ W_____, durch regelmäßiges Kiffen (verslaafd (te) raken) _____ zu _____, (bedraagt) l_____ b___ 9%. (Ter vergelijking) _____: bei Alkohol liegt sie (meer dan dubbel zo hoog) _____.

Würden Jugendliche leichter an Cannabis kommen, wenn es legal und kontrolliert vom Staat ausgegeben würde? Möglicherweise. Jedenfalls kann sich schon jetzt jeder, z.B. hier im Görlitzer Park in Berlin Kreuzberg Marihuana besorgen. (Naar de leeftijd vragen) _____ _____ die Dealer natürlich nicht. Und die Jugendlichen (in contact komen met) _____ hier _____ dem kriminellen Milieu und (harddrugs) _____ _____. Das Gras ist oft verunreinigt, zum Beispiel mit Haarspray.

Dieser Schwarzmarkt würde bei der Legalisierung vermutlich (grotendeels) z_____ _____ verschwinden. (Enorme kosten) _____ könnten so bei Polizei und Justiz (bespaard) _____ werden, denn der weitaus größte Teil ihrer Antidrogenarbeit

fällt auf Cannabisdelikte. (Sprake zijn van) _____ ein bis zwei Milliarden Euro Einsparung _____ je nach Berechnung _____.

Apropos Geld: bei einer kontrollierten Abgabe könnte der Staat sogar noch (flink verdienen) o_____ _ durch Steuern und Lizenzen für Anbau und Verkauf von Cannabis. All das wäre (goed geïnvesteerd geld) _____ a_____ _____ in Drogenprävention und Suchthilfe und könnte so vielleicht sogar (de vraag doen dalen) _____.

Der War on Drugs, der Krieg gegen Drogen jedenfalls, sei verloren, sagen sogar der Bund Deutscher Kriminalbeamter und der Schildower Kreis, eine Gruppe von Strafrechtsprofessoren. Sie (eisen een mentaliteitsverandering) _____ U_____. (Nadelen en voordelen) S_____ u_____ N_____ der Drogenpolitik müssten endlich rein wissenschaftlich überprüft werden, um ehrliche (antwoorden vinden op de vraag) _____ zu _____, ob und wie eine Legalisierung von Cannabis funktionieren könnte.

Appendix C

Sight translation (immediate post-test)

Naar schatting één Duitser op vier heeft al eens geblowd. En dat hoewel cannabis illegaal is. Ondanks strafrechtelijke vervolging, razzia's en criminalisering: het cannabisgebruik blijft stabiel. Dat bezit, teelt en handel verboden zijn, heeft blijkbaar geen effect. Moet cannabis dus gelegaliseerd worden?

Duitsland zou niet het eerste land zijn. Een voorbeeld: de Amerikaanse staat Colorado. Hier is cannabis volledig legaal. Teelt, verkoop en gebruik: alles door de overheid gecontroleerd. Het percentage van cannabisconsumenten in de bevolking is hier overigens niet of niet aanzienlijk hoger dan in andere landen waar cannabis illegaal is.

Maar toch: cannabis is een drug. De werkzame stof THC kan vooral bij jonge mensen de kans op psychose verhogen en schade aan de hersenen berokkenen. En de kans, door regelmatig blowen verslaafd te raken, bedraagt ongeveer 9%. Ter vergelijking: bij alcohol is het risico meer dan dubbel zo hoog.

Zouden jongeren makkelijker aan cannabis geraken, als het legaal en gecontroleerd door de overheid zou worden verdeeld? Misschien wel. In elk geval kan nu al iedereen, bijvoorbeeld hier in het Görlitzer Park in Berlijn Kreuzberg, marihuana kopen. Naar de leeftijd vragen de dealers natuurlijk niet. En de jongeren komen hier met het criminele milieu en harddrugs in contact. Het weed is vaak vervuild, bijvoorbeeld met haarspray.

Deze zwarte markt zou bij de legalisering vermoedelijk voor een groot deel verdwijnen. Enorme kosten zouden zo bij politie en justitie kunnen bespaard worden, want het allergrootste deel van hun antidrugsbeleid gaat naar cannabisdelicten. Er is sprake van 1 tot 2 miljard euro besparing, afhankelijk van de berekening.

A propos geld: bij een gecontroleerde verdeling zou de staat zelfs nog flink kunnen verdienen door belastingen en licenties voor teelt en verkoop van cannabis. Dat zou allemaal goed geïnvesteerd geld zijn in drugspreventie en hulp aan verslaafden en zou zo misschien zelfs de vraag kunnen doen dalen.

De War on Drugs is in elk geval verloren, zeggen zelfs de Bund Deutscher Kriminalbeamter en de Schildower Kreis, een groep van strafrechtprofessoren. Zij eisen een mentaliteitsverandering. De nadelen en voordelen van het drugsbeleid moeten eindelijk zuiver wetenschappelijk gecontroleerd worden om eerlijke antwoorden te vinden op de vraag of en hoe een legalisering van cannabis zou kunnen functioneren.

L2 reading and vocabulary development after a short Study Abroad experience

Judith Borràs
jborras@dal.udl.cat

Àngels Llanes
allanes@dal.udl.cat
Department of English and Linguistics, Universitat de Lleida, Spain

Abstract

This study explores the development of L2 reading and vocabulary as a result of a short (3-week) SA experience. Given the growing literature on Study Abroad (SA) research, this investigation attempts to shed some light on two rather under-researched areas such as L2 reading and vocabulary, and it does so with a group of adolescents (n= 52), a population often neglected by the SA research despite being one of the most common participants in SA programmes. The participants, Spanish learners of English (L2), were administered a placement test to determine their initial L2 level, a reading text (from which fluency and comprehension were calculated), the Updated Vocabulary Level Test to measure their receptive vocabulary knowledge, and a written essay to capture their productive vocabulary development in terms of lexical accuracy, fluency, density and sophistication. The results indicate that short SA experiences are positive for both reading fluency and comprehension as well as for receptive vocabulary development. Findings are not so positive in terms of productive vocabulary skills. Finally, results failed to show a direct relationship between gains in reading, and initial proficiency and vocabulary level.

Keywords: Study Abroad, reading development, vocabulary development, learning context, proficiency level

Resumen

Este estudio explora el desarrollo de las habilidades lectoras y la adquisición de vocabulario en una segunda lengua (L2) como resultado de una estancia corta (3 semanas) en el extranjero. Dado que se está dando mucha importancia a la

investigación de las estancias en el extranjero, este proyecto pretende dar envergadura a dos áreas comúnmente ignoradas como son la lectura y el vocabulario. Lo hace con un grupo de adolescentes (n= 52), un grupo poco investigado en cuanto a la investigación de las estancias en el extranjero a pesar de ser unos de los participantes más comunes en éstas. A los participantes, españoles estudiantes de inglés como L2, se les administró un test para saber su nivel inicial en la L2, un examen de lectura (para saber su fluidez y comprensión), el Updated Vocabulary Level Test para medir su vocabulario receptivo, y una redacción para capturar su vocabulario productivo (en cuanto a precisión, fluidez, densidad y complejidad). Los resultados indican que las experiencias en el extranjero, aunque sean cortas, son positivas para la fluidez y la comprensión lectora y también en cuanto a vocabulario receptivo. Por lo que hace a la mejora de vocabulario productivo, los participantes sólo mejoraron significativamente su precisión a la hora de usar la L2. Finalmente, los resultados muestran que no hay una relación directa entre la mejora de la lectura y el nivel inicial, y el vocabulario inicial de los participantes.

Palabras clave: Estancias en el Extranjero, lectura, L2, vocabulario, contexto de aprendizaje, nivel de competencia

1. Introduction

Study Abroad (SA) experiences, especially those undertaken during the summer break, are gaining much popularity today, probably because it is believed that they provide a language immersion that is essential in order to learn or improve a second or foreign language (L2) (Dewey, 2004; Kinginger, 2009).

Despite the growing importance of these short stays, little research has been carried out investigating their effects on L2 gains (Llanes, 2011), since most of the SA research has focused on the effects of semester-long SA programmes. Moreover, the vast majority of research on SA examines the effect of these stays on university students and research with adolescents is rather scarce (Evans & Fisher, 2005). Another remarkable fact is that the scant research available on the impact of a short SA experience has focused on skills such as oral fluency or pragmatics (Llanes & Muñoz, 2009; Reynolds-Case, 2013), leaving other L2 areas such as reading and lexical development in a second place. Hence, although findings seem to be positive for some L2 areas such as oral fluency, some other areas are still rather under-researched. One such neglected area is L2 reading. The very few authors who have decided to examine the effects of SA programmes on the development of reading skills have done so with participants engaged in a long (+8 weeks) SA experience (Dewey, 2004; Li, 2014; Kraut, 2016), and these studies include university students. To the authors' knowledge, there

is no study focusing exclusively on the impact of a short SA overseas experience on adolescents' L2 reading despite the fact that SA experiences lasting 2-3 weeks are the most popular among young adults and adolescents. This lack of research in reading is rather surprising considering that reading is one of the most important skills as it provides access to information (Grabe, 2009).

Another domain that plays a very important role in learning an L2 is vocabulary. As Nation (2006) and Schmitt (2008) claim, it takes much vocabulary to use a language well. This is particularly true for reading, "being the lack of lexical knowledge a major obstacle for successful comprehension even for advanced learners" (Zaytseva, 2016:45). Previous research has shown that vocabulary knowledge is related to reading development (Grabe, 2009; Li & Kirby, 2014), so examining the impact of an SA experience on lexical knowledge is necessary when analysing the SA impact on reading. Although some of the studies that have analysed the effect of an SA experience on vocabulary development report positive results (Ife, Vives, Boix, & Meara, 2000; Briggs, 2015), these studies have explored the effects of longer SA experiences (3-4 months) and only very few studies to date have examined the effect of a short SA (3-4 weeks) experience on vocabulary development. One of such studies is Llanes and Muñoz (2009), but the authors only included two vocabulary measures, namely oral lexical complexity and lexical errors. Therefore, more in-depth studies examining the impact that a short SA experience has on L2 vocabulary acquisition are needed.

All in all, given the popularity of short SA programmes and the paramount importance of L2 vocabulary and reading proficiency, the present study aims at examining whether a 3-week SA experience will lead participants to 1) improve reading comprehension and reading rate, 2) learn new L2 words and enabling them to use L2 lexicon more efficiently when producing a written task, and 3) to investigate if there is a relationship between gains in participants' L2 reading skills (if any) and their initial vocabulary knowledge and proficiency level.

2. Literature review

2.1. SA and L2 reading development

Generally, students who participate in SA experiences travel to countries where they can be fully immersed in the language in order to practise the L2 and assuming the L2 will improve. However, SA research has revealed that not all L2 areas are positively affected after an overseas experience. Although many studies conclude that SA is a good weapon towards language improvement (Juan-Garau, 2014; Michell, Tracy-Ventura, &

McManus, 2017), the same attention has not been given to all the L2 areas: There are many studies on oral skills (George, 2014; Llanes, Mora, & Serrano, 2016), but other areas such as reading require further investigation. As Kinginger (2009:61) stated "competence in reading is remarkably under-represented in the applied linguistics literature related to study abroad".

Apart from the lack of studies investigating the L2 reading development after an SA experience, another conspicuous problem is that the findings of these studies are unclear, which makes it difficult to draw decisive conclusions on whether SA has a positive impact on L2 reading. Dewey (2004), for example, analysed the L2 reading development of 30 North-American students, learners of Japanese as an L2. Half of the participants enrolled in an Intensive Immersion (IM) course in the United States, while the other half joined a SA course in Japan. Although Dewey analysed reading comprehension using three different measures (free-recall protocols, vocabulary knowledge tests and self-assessment) significant differences between the SA and the IM group were only found regarding the participants' self-assessment. More specifically, the author found that learners in the SA programme felt more confident with their reading skills after spending 11 weeks in the target country than those students who participated in the IM course. These results indicate that SA may be positive for reading skills, nonetheless, results focused exclusively on students' perceptions and there is no data providing actual information that the SA students were in fact better at reading by the time they went back to the U.S.A. A year later, Evans and Fisher (2005) examined the impact that a short (6-11 days) SA experience had on a group of English secondary school students, learners of French as an L2 (n= 68). The authors attempted to provide evidence of the impact that participating in a school exchange programme had on the participants' French proficiency. They concluded that such a short period of time may provide significant gains in listening and writing skills but no meaningful influence was found in terms of L2 reading or speaking, suggesting that short SA programmes do not provide enough practise for learners to improve their L2 reading skills. More promising results come from Li (2014), who investigated the L2 reading development of six groups of North-American undergraduates, learners of Chinese (n= 73), over an 8-week SA experience in China. Li's groups consisted of a beginner group at home (AH) and another one SA, an intermediate group AH and their SA counterparts, and an advanced group AH and their SA counterparts. The author examined the participants' reading comprehension skills and their use of strategies when reading in Chinese. Li found that, generally, the SA groups outscored those AH in all the measures analysed. Hence, the author concluded that SA was positive for both L2 reading comprehension and strategy use, with intermediate students showing the greatest gains. In a more recent study, Kraut (2017) examined a group of 16 Arabic and Chinese students, learners of L2 English, who were enrolled in an 8-week long

Intensive English Programme (IEP) in the U.S.A. The author investigated the effects of SA on lexical inferencing abilities, vocabulary breadth, reading comprehension and reading speed, and participants' attitudes towards reading. Kraut's results show a positive picture for participants' reading speed, willingness to read out of curiosity and students' perceived self-efficacy. It must be highlighted, though, that participants in Kraut's study participated in an IEP programme, which involves more hours of instruction per week (Kraut, 2017) than traditional SA programmes. Hence, the combination of the intensity of instruction and SA indicate that IEPs may provide greater chances for learning to occur. All in all, the impact that SA programmes have on L2 reading skills is still somewhat blurred and no conclusions can be drawn to determine whether living in the target community aids the development of the reading skill. Moreover, most of the studies which have investigated this skill have a duration of +8 weeks and are conducted with undergraduate students. There is only one study to date analysing the effects that a shorter SA programme may have on the L2 reading development with group of adolescents (Evans & Fisher, 2005) and it does so in a very broad manner: it not only focused on reading skills but on L2 development in general, including listening, reading, writing and speaking). Therefore, it only provides an overview of the development of students in each area and, consequently, more studies investigating if short SA experiences influence L2 reading positively are needed.

2.2. SA and L2 vocabulary development

Several authors have claimed that readers need a large amount of vocabulary in order to understand a text well (Hu & Nation, 2000; Nation, 2006; Schmitt, 2008). Not knowing the meaning of certain words in a text can be a major obstacle for reading fluency since readers will not be able to read at their normal speed if they have to stop because of their lack of vocabulary knowledge (Beglar, Hunt, & Kite, 2012; Huffmann, 2014). Since previous studies on SA have found that vocabulary is one of the areas positively influenced after an SA, it is plausible that this improvement in L2 vocabulary (if any) could be reflected in the participants' reading skills.

A growing number of studies have examined the impact of SA on L2 lexical knowledge (Ife, Vives-Boix, & Meara, 2000; Foster, 2009; Pérez-Vidal & Barquin 2014; Zaytseva 2016). As an illustration, Barquin (2012) analysed the written vocabulary development of a group of 30 Catalan/Spanish bilinguals, learners of English as an L2 first AH and, then, abroad. She concluded that the period abroad was significantly favourable for productive vocabulary development. Another relevant study on SA and vocabulary is that by Briggs (2015), in which the author explored if out-of-class activities affected gains in receptive and productive vocabulary. Participants in her study (n= 241) had different first languages and studied English in the UK for different

periods of time (short = 6-10 weeks, medium = 11-15 weeks, long = 16-20 weeks). Briggs concluded that SA led to gains in vocabulary breadth although she found that these gains were not related to place or out-of-class contact. The author also claimed that length of stay was an important predictor of language gains: "the longer a learner spends in the study abroad context, the higher their gains" (2015:137) with participants staying for 6-10 weeks improving their receptive vocabulary significantly but failing to show gains regarding productive vocabulary. Another study that examined the impact of an SA experience on vocabulary is Tracy-Ventura (2017), who explored the lexical sophistication of a group of 27 L1-English learners of Spanish in Spain and Mexico. The author found that participants improved significantly their knowledge of low frequency words after their 9-month-long SA experience. Moreover, statistically significant changes were also found with regards to participants' actual use of these words indicating that participants not only learnt low frequency words, but they also used them more. Altogether, this study provides positive evidence of SA and lexical knowledge. Noticeably, it can be stated that SA is beneficial for vocabulary development as there are many different examples which prove it. Nonetheless, as in the case of reading, none of them investigates the vocabulary development of teenagers as a result of a short overseas experience. Moreover, the few studies which have examined the effects of a short SA experience on L2 vocabulary (Llanes, 2012; Serrano, Llanes, & Tragant, 2016) do not focus exclusively on vocabulary (they examine different L2 domains) and, hence, no study has included a relatively wide range of measures of L2 receptive and productive vocabulary, which would provide more robust results.

There is evidence that short (≤ 5 weeks) SA experiences have a positive impact on several L2 areas, such as oral skills (Llanes & Muñoz, 2009), pragmatic development (Reynolds-Case, 2013), listening comprehension (Cubillos, Chieffo, & Fan, 2008; Rodrigo, 2011), pronunciation (Llanes, Mora, & Serrano, 2016), and writing skills (Evans & Fisher, 2005). However, there is no single study, to the authors' knowledge, that looks at the impact that such a short SA experience has on L2 vocabulary, and there is only one that includes reading development (Evans & Fisher, 2005), although learning vocabulary is usually among the short-SA participants' goals (Allen, 2010) and reading is one of the most important skills (Rasinski, 2003). There are reasons to think that both skills would improve after an SA experience. One of such reasons is the intensive exposure that SA participants experience while overseas, and previous research has found that intensive experiences are conducive to L2 gains (Muñoz, 2012). Another reason for potential gains in vocabulary is the ample opportunities for practise, and practise has been found to be a key factor when improving the L2 (DeKeyser, 2007). However, the impact that the SA context has on L2 reading development is especially interesting because this skill is hardly ever (extensively) practised while abroad, and simply being in the target country may not guarantee that

students improve this skill. It is plausible to think that SA participants have chances of improving their L2 vocabulary repertoire because, apart from the amount of exposure and practise that characterize the SA context, participants use the L2 for real purposes, and this may result in an improvement in vocabulary. It is also reasonable to think that if there is an improvement in vocabulary, this improvement can positively affect reading skills, at least as far as reading fluency and comprehension are concerned. However, whether three weeks is enough for gains in L2 reading and vocabulary to occur is unknown. Previous studies examining the impact of a specific treatment on L2 vocabulary development (not necessarily in a SA context) have found gains in as short a period of time as three to six days (Beck & McKeown, 2007), two weeks (Kaivanpanah & Rahimi, 2017) or one or two semesters (Ife, Vives, Boix, & Meara 2000). Hence, it is likely that there is improvement in vocabulary after three weeks overseas. Nonetheless, the dearth of research in the area makes it difficult to draw robust conclusions on whether short SA experiences are positive for participants' L2 reading and vocabulary development.

Thus, in order to fill these existing gaps in the literature, the current study addresses the following questions:

1. Does a 3-week SA experience have a positive impact on L2 reading development in terms of comprehension and fluency?

2. Does a 3-week SA experience have a positive impact on L2 receptive vocabulary development (measured through the Updated VLT) and productive vocabulary development (measured through lexical fluency, lexical accuracy, lexical density and lexical sophistication)?

3. Are gains in L2 reading (in terms of comprehension and fluency) related to initial L2 vocabulary knowledge and proficiency level?

3. Method

3.1. Participants

Fifty-two teenagers (n= 25 males, n= 27 females), learners of English as an L2 participated in the present study. Their ages ranged between 12 and 17 years old (M= 15,35) and all came from different regions of Spain but one, who came from Andorra. All the participants started learning English at primary school and their levels by the time they started the SA experience ranged from A2 to B2 (A2 n= 31, B1 n= 17 and B2 n= 4). The students enrolled in a 3-week summer programme in Ireland which fostered

communication in English. All the participants lived with a host family and attended classes Monday through Friday for 4 hours per day during the mornings. The teachers were all native English speakers and the activities carried out in the classrooms were very dynamic and consisted in playing games, performing some debates, and completing some worksheets, among others. When classes finished, students were given one hour to have lunch and then they spent two more hours at the school carrying out different activities with the teachers: Irish workshops, Irish sports, arts and crafts, or visiting some tourist places. After the afternoon activities students spent the evenings with their host families, with whom they performed different activities.

3.2. Instruments

The Oxford Quick Placement Test (OPT) was used to examine participants' overall L2 proficiency. This test has been widely used by researchers (Llanes et al., 2016) and it has proved to be a reliable instrument to examine L2 proficiency. The test contains two sections, but only the "Use of English" one was used during the data collection. This part consisted of 60 multiple choice items with three or four possible answers, and one point was given per correct item.

Two texts of different proficiency levels were chosen to measure participants' reading fluency and comprehension. Following Kraut (2017), both texts were taken from the *Reading Explorer*, a book that contains texts of different lengths and diverse topics. The texts are followed by some comprehension questions. To make sure that the texts targeted different L2 levels and that they were suitable for the participants, their readability index was also calculated. The first was taken from the *Reading Explorer 2*, which contains texts of an A2 level according to the Common European Framework of Reference for Languages (CEFR). The readability test showed a 72.3 in Flesh reading ease, which means that the difficulty of the text was equivalent to texts generally used during the first year of secondary education. The second text was taken from the *Reading Explorer 3*, which contains texts of a B1 CEFR level. This B1 text obtained a 52.8 in Flesh reading ease, indicating that the difficulty of the text was similar to texts read during the last years of secondary education and the A levels. Each text was followed by a set of multiple choice questions. Although most of the questions came from the Reading Explorer book, some were slightly changed and some others were added by the authors (the added questions were previously piloted). After determining the participants' initial proficiency level, it was decided that only the data of the A2 text would be considered for the present study given that previous research shows that one requirement to measure reading rates is text suitability: the text has to be well within the students' capability (Carver, 1990; Huffman, 2014).

Two instruments were used to gauge participants' lexical knowledge: the Updated Vocabulary Levels Test (VLT) (Webb, Sasao, & Balance, 2017) and a writing task. The Updated VLT measures receptive vocabulary and has been widely used in the research of L2 lexical knowledge (Briggs, 2015). This test is divided into five levels (K1, K2, K3, K4 and K5) and words are provided in sets of six words with three possible definitions; participants have to match the items with their proper definition (with the three remaining words as distractors). There are 30 definitions per level and five levels in total, which amounts to 150 definitions. The writing task measured productive vocabulary. Since researchers who have analysed vocabulary claim that much data can be gathered from a written text (Zaytseva, 2016; Tracy-Ventura, 2017) it was believed that this task would be a reliable tool to gather data for lexical knowledge and lexical error analysis. Participants were allotted fifteen minutes to write a text entitled "My life: past, present and future expectations" (Llanes & Serrano, 2014) and no specific word limit was required. It must be highlighted that productive vocabulary was measured through a written test, but writing skills development (i.e. grammar, complexity or syntax) was not the focus of the present study.

In order to triangulate the data, the researcher who conducted the pre- and post- tests also spent some time observing the participants in their classes, afternoon activities, weekends and with their families. Moreover, the researcher also interviewed, informally, some of the teachers and students in order to have more information on the overall learning experience.

3.3. Procedure

This study has a pre- and post-test design. The tests were exactly the same for the pre- and post-test. However, participants were not informed about the focus of the research and that there would be a post-test so that this could not affect the results. Both tests were administered in Ireland: the pre-test was administered on the third day into the SA programme, and the post-test two days before the students' departure to Spain. Thus, the stay lasted 3 weeks, but the time elapsed between the pre- and post-test was 15 days. The procedure was the same at both times, the only difference being that in the pre-test students were asked to fill out a personal information sheet before administering the tests. First, the two texts were administered. A chronometer was projected on the front screen of a digital board so that everyone could see it. Then, the participants were given a text to read and, once they had finished, they were asked to look at the chronometer and write down the time they had taken to read it (minutes and seconds). Participants were informed that they could only read the text once, that they would not have the text when answering the comprehension questions, and that they were asked to read the text at their normal speed. In order to counterbalance any task-effects, half of the

participants started reading the A2 text and the other half the B1 text (the students that started reading the A2 text in the pre-test also did so in the post-test, and the same was true for the B1 text). Afterwards, the texts were collected and the comprehension question sheet was administered. The students were not given any specific time to answer the questions and the researchers waited until all of them had finished to start with the next exercise, the written essay. Participants were asked to write a composition with the title "My life: past, present and future expectations" and they were allotted a total of 15 minutes. After the writing task, the students were asked to fulfil the Updated VLT, for which they were given a maximum of 20 minutes. Finally, participants were administered the OPT and they were given 30 minutes. Once they had finished the OPT, the data collection was completed and students were given permission to leave the room. The whole data collection procedure took approximately 90 minutes.

3.4. Measures

Reading fluency was measured through words per minute (WPM). The following formula was used to compute WPM: 504 (which is the number of words the text had) x 60 seconds/ total time in seconds students took to read the text. Concerning reading comprehension, the number of correct answers out of 13 was used to measure text comprehension.

As for receptive vocabulary, the updated VLT provided information about the participants' lexical knowledge in terms of receptive vocabulary knowledge. The test included 150 target words, so the raw score out of 150 was calculated. With regards to productive vocabulary, the measures in the present study were adopted following Zaytseva (2016). The Computerized Language Analysis (CLAN) software was used to measure lexical fluency and accuracy. The programme provided the total number of types, tokens and lexical errors the students had made in the texts and with this information the participants' lexical fluency and accuracy were calculated. Lexical fluency was measured through tokens (i.e. total number of words used to write the text) and lexical accuracy was measured by counting the amount of lexical errors in the text and dividing them by the total number of tokens. In order to calculate the participants' lexical density and sophistication the online tool VocabProfile was used. Lexical density consisted of the percentage of content words (nouns, verbs, adjectives and adverbs) in the text. To measure lexical sophistication, the number of "rare" words that participants used within the text (words from the 6000 frequency level or above) was used to see whether they used words from a higher frequency level (i.e. more sophisticated) in the post-test writings. Finally, the CEFR level of the participants was calculated by the total number of points they obtained from the OPT (see Table 1 for a summary of the measures used).

Higher values in the post-test were expected for all the measures except for lexical accuracy, since lower values in this measure would mean that participants had made fewer lexical errors in their post-writings. In terms of lexical sophistication, a lower percentage of words from the 1000 frequency level would also indicate that participants' texts were more sophisticated by the time they left the host country.

Table 1: Summary of measures

Domain	**Measure-Test**	**Formula**
General proficiency	OPT	Raw score /60
Reading fluency	WPM - reading text	504*60/Total seconds taken to read the text
Reading comprehension	Comprehension - reading text	Raw scores /13
Receptive vocabulary	VLT	Raw scores /150
Productive vocabulary	Lexical fluency - written essay	Total number of tokens
	Lexical accuracy - written essay	Number of lexical errors/ tokens
	Lexical density - written essay	Number of content words/tokens
	Lexical sophistication - written essay	Comparison of the percentages from words of different frequency levels

4. Results

When checking the normality of the data, most of the measures violated the assumption of normality. Therefore, non-parametrical tests were run to answer the first and second research questions. However, before running any tests, the data were coded independently by the first author of the present study and by another expert to ensure inter-rater reliability. The two encodings were compared and they reached an agreement of 92.31%. Intra-rater reliability was also calculated, and the level of agreement was exactly the same (92.31%). Inter- and intra-rater reliability were only calculated for lexical errors given that it was the only measure that could vary depending on the rater.

The first research question asked whether a 3-week SA experience had a positive impact on L2 reading development in terms of comprehension and fluency. In order to answer this question, a Wilcoxon signed-rank test was run to observe if there were significant differences between the scores in the pre-test and those in the post-test. As shown in Table 2 below, participants improved on both measures of reading (WPM and comprehension) from the pre- to the post-test, and the Wilcoxon test indicated that improvement was statistically significant for both: WPM (Z= -3.201, p= .001), and comprehension (Z= -2.902, p= .004). The effect sizes for WPM and Comprehension were d= 0.492, d= 0.55, respectively, which indicate that this difference was medium (Cohen, 1988).

Table 2: Descriptive statistics

Measure	**Pre-test**		**Post-test**	
	M	SD	M	SD
OPT	29.25	6.90	30.65	7.05
WPM	118.65	33.29	134.60	32.73
Comprehension	6.73	1.87	7.54	1.83
VLT raw score	102.19	28.06	112.56	26.66
Lexical fluency	127.08	39.13	128.02	40.268
Lexical accuracy	.037	.024	.013	.014
Lexical density	.45	.043	.45	.045
1000 Frequency Level	90.69	4.25	91.71	3.66
>5000 Frequency Level	2.45	1.34	1.95	.75

The second research question addressed whether a 3-week SA experience had a positive impact on L2 lexical knowledge. Table 2 shows that participants improved their receptive vocabulary (VLT) and two measures of productive vocabulary (lexical fluency and accuracy). However, there was no improvement regarding lexical density and lexical sophistication. Again, a Wilcoxon signed-rank test was employed to see if these pre- to post-test differences were significant and it was found that their receptive vocabulary significantly improved from the pre- to post-test (Z= -5.132, p= .000), but only one measure of productive vocabulary improved significantly, namely lexical accuracy (Z= -5.689, p= .000). Therefore, the data show that participants learnt new words and made significantly fewer lexical errors in their post-test essays. However,

although there seems to be a slight tendency for learners to produce longer essays, there was no significant difference in terms of lexical fluency between the essays before and after the stay. The effect sizes indicated that the difference for the two measures that changed significantly was medium for receptive vocabulary (d= 0.365) and large for lexical accuracy (d= 1.22).

The last research question asked whether gains in reading (both in terms of comprehension and fluency) were related to initial L2 proficiency and initial vocabulary knowledge. In order to answer this question, bivariate correlations were run between initial L2 level (OPT pre-test scores), initial L2 vocabulary level (pre-test raw scores of VLT) and gains in reading comprehension and fluency (to calculate the gains in comprehension and fluency, the score in these measures in the post-test was subtracted from the score in these measures in the pre-test). No significant correlations were found between any of the independent variables and gains in reading.

5. Discussion

The aim of this study was to document the impact of a 3-week SA experience on adolescents' L2 reading and vocabulary development as well as to examine to what extent gains in reading and vocabulary (if any) were related to L2 initial proficiency level and initial lexical knowledge. It was found that participants improved the two measures of L2 reading significantly, namely fluency and comprehension. Hence, the answer to our first research question is affirmative. Although the participants did not explicitly engage (much) in reading activities, they were massively exposed to English. This exposure, although not being through reading explicitly, may have been helpful for the students' reading development. As Gautier and Chevrot (2015:169) claimed, "Learning in contexts where the target language is used is considered particularly beneficial because such an environment should provide access to language that is ample in quantity and diverse in quality". Therefore, this intensive immersion in the L2 may have fostered the improvement of the participants' L2 reading skills (Muñoz, 2012).

These results are in line with Kraut (2017), who found that an 8-week-long SA experience had a positive impact on L2 reading skills. However, the present study shows that even shorter SA programmes (3 weeks) impact positively L2 reading development in terms of comprehension and fluency. The positive outcomes found in the present study could also be explained because of the participants' proficiency level. Previous studies show that participants with an advanced proficiency level do not progress as much as participants with a lower proficiency level, suggesting that the lower the participants' initial proficiency level, the more chances they have to improve it (Juan-

Garau, 2014). Since most of the participants in the present study had an A2 or a B1 proficiency level, there was plenty of room for improvement and the measures used may have captured these gains. Another possible explanation to the positive outcomes in terms of reading is the intensity of instruction and/or the nature of SA experiences, which is a combination of formal in-class learning with informal out-of-class learning (afternoon activities and interaction with the members of the family).

The second research question asked whether a 3-week SA experience had a positive impact on L2 receptive and productive vocabulary. It was found that participants improved their receptive vocabulary significantly from the pre- to the post-test, and that they improved significantly one out of the four productive vocabulary measures, namely lexical accuracy. Therefore, results suggest that short SA experiences are positive in terms of learning new words and reducing lexical errors, but results also suggest that three weeks are not enough for participants to improve their lexical fluency, density and sophistication significantly. Hence, findings in the present study suggest that 3-week-long stays are not enough for learners to write significantly longer, denser and more sophisticated essays. All in all, the answer to the second research questions is affirmative in terms of receptive vocabulary, but its effects on productive vocabulary are not so promising. The reason for finding significant improvement in some vocabulary measures (receptive vocabulary and lexical accuracy) and not in others (lexical fluency, density and sophistication) might be the short time elapsed between the pre- and the post-test. It is possible that measures such as receptive vocabulary and lexical accuracy were more susceptible to gauging gains than other measures such as lexical density or sophistication, which might need more time to develop (in line with Briggs, 2015). With regard to receptive vocabulary, the results in the present study support Kraut (2017), who found similar results with her participants after an 8-week SA. In fact, a closer look at the VLT revealed that some participants learnt words such as 'sheet', 'alley' or 'forbid', which are likely to be encountered in an SA experience. Likewise, the lexical accuracy of the participants was also examined more closely. Table 3 includes excerpts of productions from the same participants in the pre and the post-test, where examples of lexical accuracy development can be clearly appreciated. Some of the lexical accuracy improvements SA students made involve a) appropriate distinction between words earlier treated as homophones (i.e. leave/live), b) reduction of made-up words (*tought/play), and c) more accurate lexical choice (travel/trip).

Table 3. Examples of lexical accuracy from the same students´ pre- to post-test

Student	Pre-test	Post-test
Student A	I **leave** in Palma. Before I **leaved** in Madrid.	I **live** in Palma. Before I **lived** in Madrid.
Student B	I **tought** the flut.	I **play** the flute.
Student C	I am the **unic** member of my family who…	I am the **only** member of my family who…
Student D	I am going to **curse** 3rd of ESO.	I am going to **do** 3rd of ESO.
Student E	I was waiting for this **travel**.	I loved this **trip**.

Other previous studies that have found gains in some of the productive vocabulary measures analysed in the present study are those by Pérez-Vidal and Barquin (2012) and Zaytseva (2016). These studies provide evidence that measures such as lexical fluency or density can be improved during an SA experience. Notwithstanding, there are three important differences between these studies and the present one: the duration of the programmes, the participants' age and their initial proficiency level. Participants in these studies engaged in a 3-month SA experience and they were university students with a higher initial L2 level. Previous research in the field of vocabulary has found that high gainers on comprehension made significantly larger gains in vocabulary (Shany & Biemiller, 2010). Hence, it is plausible that for productive vocabulary learning, a higher initial L2 level could have had a more positive impact. Given that reading and vocabulary are two different skills and the nature of the tests administered was also different, it is possible that participants with a lower L2 level improved reading comprehension and fluency more than vocabulary, whereas participants with a higher L2 level improved their vocabulary knowledge more. A tentative explanation for this latter hypothesis is that participants with a higher L2 level could possibly allocate their attention to other aspects of the L2 such as vocabulary. Finally, in terms of lexical sophistication, findings in previous studies seem to suggest that long periods abroad are needed in order for participants to show significant changes with regard to the measure. Zaytseva (2016), for example, found no significant difference for the size of gains in lexical sophistication after participants in her study had been abroad for three

months. However, Tracy-Ventura (2017) found that after nine months abroad not only did her participants improve their knowledge of low frequency words, but they also started using these words more in their written and oral discourses. These results provide further evidence that measures such as lexical sophistication need more time to develop than others such as lexical accuracy.

The third research question enquired whether L2 reading gains were related to initial proficiency level and initial lexical knowledge, and no significant correlation was found between these variables. Hence, the present study failed to find a direct relationship between initial vocabulary knowledge and proficiency level, and gains in reading, which suggests that all participants held the same chances to improve their reading skills regardless of their initial vocabulary knowledge and proficiency level. This finding does not support Nation (1993) and Li and Kirby (2014), who reported a relationship between vocabulary knowledge and reading skills which implied that being proficient in one of the skills was the factor that helped the most when attempting to improve the other. The present study suggests that SA adolescents improve their reading comprehension and fluency regardless of their initial L2 grammar or vocabulary knowledge. This finding indicates that a short SA is positive for teenagers' reading skills and not only for those who have a specific proficiency level or vocabulary knowledge. Nonetheless, it must be highlighted that the proficiency level of participants in the present study ranged mostly from A2 to B1 and that different results could have emerged if participants had a more advanced level. Previous research has exhibited a certain contradiction regarding the precise level that participants in SA need to hold before starting their stays (the threshold level). Some scholars such as Llanes and Muñoz (2009) and Juan-Garau (2014) state that 'the lower, the better'. These researchers argue that participants with lower L2 skills have more room to learn and, hence, their improvement is easier to detect. On the other hand, other authors such as DeKeyser (2007) and George (2014) claim that participants should possess a good command on the L2 before starting their stay in order to make the most of the SA. The present study is taking a central position between these two groups of researchers since it shows that the initial vocabulary knowledge and the proficiency level of the participants are not related to gains in reading skills and that, therefore, the threshold level is not a variable that has a strong effect on reading and vocabulary gains. It is possible that factors such as motivation to learn, willingness to take advantage of the SA, living with a host family, the type of programme or the personality of the participants, among others, are the constituents which have helped participants in the present study to improve their reading comprehension and fluency (see also Saito et al., 2018).

6. Conclusion

No previous studies exist on the L2 reading and vocabulary development of a group of teenagers engaged in a short SA experience. It was found that participants improved their L2 reading comprehension and fluency significantly, showing greater understanding of texts and a faster reading rate after the 3-week stay abroad. Moreover, the results in the present study show that participants significantly improved their receptive vocabulary knowledge, which indicates that they learnt a significant number of new words during their time abroad, and their lexical accuracy, which indicates that 3-week-long SA help learners use L2 words in a more accurate way. No significant differences were found concerning the remaining three productive measures examined (lexical fluency, lexical sophistication and lexical density), suggesting that 3-week SA programmes are not enough for students to write longer, denser and more sophisticated texts. Although the findings of the present study suggest that some measures of productive vocabulary (fluency, density and sophistication) need more time to develop, results also suggest that short SA programmes can have a positive feedback on L2 reading and vocabulary development.

However, this investigation has some limitations. One of the limitations is that it does not include a control group learning the L2 at home (AH). Although the inclusion of control groups has been questioned in the SA literature because of the large differences between the participants engaging in an overseas experience and those remaining AH, it would be interesting that further research include a comparison group AH engaged in a course focused on reading. This way, the effects of exposure to the L2 could be compared to the effects of reading practise and more robust conclusions could be drawn. Another limitation is the lack of more precise information on the amount of input and practise experienced. In other words, it would be interesting to know the amount of time participants spent speaking, reading or writing in English and see if amount of time spent practicing the L2 explained gains in reading or vocabulary. Finally, although the researchers made an effort to avoid task-repetition effects (the inclusion of two reading tests, not informing the participants that there would be a post-test, and administering the reading texts at the very beginning of the data collection), it is still possible that using the same instruments in the pre- and post-test might have influenced the results. Despite these limitations, the present study is a relevant contribution to the field of SA as it sheds light on the impact that a 3-week SA experience has on the L2 reading and vocabulary development of a group of adolescents.

Given that the present study shows that a 3-week SA experience has a positive impact on the L2 reading development and on some measures of vocabulary development, short SA programmes should be promoted among teenagers. Moreover, considering

that teenagers who undertake summer SA experiences of this type usually do so via private institutions, these SA experiences should be made affordable to teenagers so that they will be able to participate and boost their L2 skills.

Acknowledgements

This study was supported by grant 2017SGR1522. We would like to thank Ferran C. for letting us conduct the data collection at one of his summer courses abroad and for his efforts to make the process as easy as possible. Our thanks are extended to Suzanne and Billy, teachers of the Irish language school, for letting us conduct the data collection and for helping us throughout the process. We would also like to thank the leaders for their comments and help during the data collection. Finally, we would like to thank the editor and the reviewers of VIAL for their suggestions on this article.

7. References

Allen, H.W. (2010). Language-learning motivation during short-term study abroad: An activity theory perspective. *Foreign Language Annals,* 43(1), 27-49.

Barquin, E. L. (2012). *Writing development in a study abroad context* (Unpublished doctoral dissertation). Universitat Pompeu Fabra, Barcelona.

Beck, I.L., & McKeown, M.G. (2007). Increasing low-income children's oral vocabulary repertoires through rich and focused instruction. *The Elementary School Journal*, 107(3), 251-271.

Beglar, D., Hunt, A., & Kite, Y. (2012). The effect of pleasure reading on Japanese university EFL learners' reading rates. *Language Learning,* 62(3), 665-70.

Briggs, J. G. (2015). Out-of-class language contact and vocabulary gain in a study abroad context. *System,* 53, 129-140.

Carver, R. P. (1990). *Reading rate: A review of research and theory*. New York: Academic Press.

Cohen, J. (1988). *Statistical power of analysis for the behavioural sciences.* Lawrence Erlbaum Associates, Hillsdale.

Cubillos, J.H., Chieffo, L., & Fan, C. (2008). The impact of short-term study abroad programs on L2 listening comprehension skills. *Foreign Language Annals*, 41(1), 157-185.

Dekeyser, R. (2007). *Practice in a second language: Perspectives from applied linguistics and cognitive psychology*. Cambridge University Press.

Dewey, D. (2004). A comparison of reading development by learners of Japanese in intensive and domestic immersion and study abroad contexts. *Studies in Second Language Acquisition,* 26, 303-327.

Evans, M., & Fisher, L. (2005). Measuring gains in pupils' foreign language competence as a result of participation in a school exchange visit: the case of Y9 pupils at three comprehensive schools in the UK. *Language Teaching Research,* 9(2), 173–192.

Foster, P. (2009). Lexical diversity and native-like selection: the bonus of studying abroad. In B. Richards, H.M., Daller, D.D. Malver, P. Meara, J. Milton & J. Treffers-Daller (Eds.), *Vocabulary Studies in First and Second Language Acquisition*. (pp 91-106) Hampshire: Palgrave Macmillan.

Gautier, R., & Chevrot, J.P. (2015). Social networks and acquisition of sociolinguistic variation in a study abroad context: A preliminary study. In R. Mitchell, N. Tracy-Ventura & K. McManus (Eds.), *Social Interaction, Identity and Language Learning during Residence Abroad*. (pp 169-184) EuroSLA Monographs Series, 4. Amsterdam: The European Second Language Association.

George, A. (2014). Study abroad in central Spain: The development of regional phonological features. *Foreign Language Annals,* 47(1), 97–114.

Grabe, W. (2009). *Reading in a second language: Moving from theory to practise*. Cambridge Applied Linguistics.

Hu, M., & Nation, I.S.P. (2000). Unknown vocabulary density and reading comprehension. *Reading in a Foreign Language,* 13(1), 403–430.

Huffman, J. (2014). Reading rate gains during a one-semester extensive reading course. *Reading in a Foreign Language,* 26(2), 17-33.

Ife, A., VivesBoix, G., & Meara, P. (2000). The impact of study abroad on the vocabulary development of different proficiency groups. *Spanish Applied Linguistics,* 4(1), 55-84.

Juan-Garau, M. (2014). Oral accuracy growth after formal instruction and study abroad: Onset level, contact factors and long-term effects. In C. Pérez-Vidal (Ed.), *Language Acquisition in Study Abroad and Formal Instruction Contexts*. (pp 87-109) Amsterdam: John Benjamins Publishing Co.

Kaivanpanah, S. & Rahimi, N. (2017). The effect of contextual clues and topic familiarity on L2 lexical inferencing and retention. *Porta Linguarum,* 27, 47-61.

Kinginger, C. (2009). *Language learning and study abroad: A critical reading of research.* Houndsmills, Basingstoke, UK: Palgrave/Macmillan

Kraut, R. E. (2017). The development of L2 reading skills: A case study from an eight week intensive English program course. *Dialogues: An interdisciplinary Journal of English Language Teaching and Research,* 1(1), 25-43.

Li, L. (2014). Language proficiency, reading development, and learning context. *Frontiers: The Interdisciplinary Journal of Study Abroad,* 24, 73-92.

Li, M., & Kirby, J.R. (2014). The effects of vocabulary breadth and depth on English reading. *Applied Linguistics,* 36(5), 611–634.

Llanes, À. (2011). The many faces of study abroad: An update on the research on L2 gains emerged during a study abroad experience. *International Journal of Multilingualism,* 8(3), 189-215.

Llanes, À. (2012). The short- and long-term effects of a short study abroad experience: The case of children. *System,* 40, 179-190.

Llanes, À., Arnó, E., & Mancho-Barés, G. (2016). Is a semester abroad in a non-English speaking country beneficial for the improvement of English? The case of Erasmus students using English as a lingua franca. Special issue of *The Language Learning Journal*, 44(3), 292-303.

Llanes, À., Mora, J.C., & Serrano, R. (2016). Differential effects of SA and intensive AH courses on teenagers. *International Journal of Applied Linguistics,* 27(2), 470-490.

Llanes, À., & Muñoz, C. (2009). A short stay abroad: Does it make a difference? *System,* 37(3), 353-365.

Llanes, À., & Serrano, R. (2014). The effectiveness of classroom instruction 'at home' vs. study abroad for learners of English as a foreign language attending primary, secondary school and university. *The Language Learning Journal,* 45(4), 434-446.

Mitchell, R., Tracy-Ventura., N., & McManus, K. (2017). *Anglophone students abroad: Identity, social relationships and language learning.* New York: Routledge.

Muñoz, C. (2012). *Intensive exposure experiences in second language learning.* Multilingual Matters: Bristol (United Kingdom)

Nation, I.S.P. (2006). How large a vocabulary is needed for reading and listening? *The Canadian Modern Language Review/La Revue canadienne des langues vivantes,* 63(1), 59–82.

Pérez-Vidal, C., & Barquin, E. (2014). Comparing progress in writing after formal instruction and study abroad. In. C. Pérez-Vidal (Ed.), *Language Acquisition in Study Abroad and Formal Instruction Contexts* (pp. 217-234). Amsterdam: John Benjamins Publishing Co.

Rasinski, T.V. (2003). *The fluent reader: Oral reading strategies for building word recognition, fluency, and comprehension*. Better World Books Ltd (Dunfermline, United Kingdom)

Reynolds-Case, A. (2013). The value of short-term study abroad: An increase in students' cultural and pragmatic competency. *Foreign Language Annals,* 46(2), 311–322.

Rodrigo, V. (2011). Contextos de instrucción y su efecto en la comprensión auditiva y los juicios gramaticales: ¿Son comparables cinco semanas en el extranjero a un semestre en casa? *Hispania,* 94(3), 502–513.

Saito, K., Dewaele, J.M, Abe, M., & In'nami, Y. (2018). Motivation, emotion, learning experience and second language comprehensibility development in classroom settings. *Language learning.* 68(3), 709-743.

Schmitt, N. (2008). Review article: Instructed second language vocabulary learning. *Language Teaching Research,* 12(3), 329–363.

Serrano, R., Llanes, À., & Tragant, E. (2012). Examining L2 development in two short-term intensive programs for teenagers: Study abroad vs. "at home". *System,* 57, 43-54.

Shany, M., & Biemiller, A. (2010). Individual differences in reading comprehension gains from assisted reading practice: Pre-existing conditions, vocabulary acquisition, and amounts of practice. *Reading and Writing: An Interdisciplinary Journal*, 23(9), 1071-1083.

Tracy-Ventura, N. (2017). Combining corpora and experimental data to investigate language learning during residence abroad: A study of lexical sophistication. *System*, 71, 35-45.

Webb, S., Sasao, Y., & Balance, O. (2017). The updated vocabulary levels test: Developing and validating two new forms of the VLT. *ITL International Journal of Applied Linguistics,* 168(1), 33-69.

Zaytseva, V. (2016). *Vocabulary acquisition in study abroad and formal instruction: An investigation on oral and written lexical development* (Unpublished doctoral dissertation). Universitat Pompeu Fabra, Barcelona.

Does younger mean better? Age of onset, learning rate and short-term L2 proficiency in young Danish learners of English

Teresa Cadierno
Department of Language and Communication, University of Southern Denmark
cadierno@sdu.dk

Mikkel Hansen
Department of Psychology, University of Paris 8
mikkel.hansen@iedparis8.net

Jørgen T. Lauridsen
Department of Business and Economics, University of Southern Denmark
jtl@sam.sdu.dk

Søren W. Eskildsen
Department of Design and Communication, University of Southern Denmak
swe@sdu.dk

Katalin Fenyvesi
Department of Language and Communication, University of Southern Denmark
katalin@sdu.dk

Signe Hannibal Jensen
Department of Language and Communication, University of Southern Denmark
sihaje@sdu.dk

Maria Vanessa aus der Wieschen
Department of Design and Communication, University of Southern Denmark
mwies@sdu.dk

Abstract

This paper reports the results of a semi-longitudinal study investigating the role of age of onset in early foreign language (English) learning. We compared two groups of Danish school children (N = 276) who, following an educational reform in 2014, started their first English classes the same year but at different ages. One group (the early starters) was introduced to English in the 1st grade (age 7-8) and the other

group (the late starters) in the 3rd grade (age 9-10). Children's receptive vocabulary, receptive grammar, and phonetic discrimination skills were followed for three years, allowing comparisons across groups and time and tracking learning rate and short-term proficiency (after one and two years of instruction). Results showed that the late starters outperformed the early starters in most tests. With respect to learning rate, the tests also revealed that the late starters had an advantage in the receptive grammar test, whereas the phonetic discrimination test showed a more diffuse picture with the late starters seemingly halting in development and the early starters advancing. The results also showed gender differences with boys achieving a higher level of proficiency and exhibiting a faster learning rate than girls. The pedagogical implications of the results are discussed.

Keywords: Early foreign language learning; receptive vocabulary; receptive grammar; phonetic discrimination; semi-longitudinal study

Resumen

Este trabajo presenta los resultados de un estudio semi-longitudinal que investigó el papel de la edad en el aprendizaje del inglés como lengua extranjera. Se compararon dos grupos de niños daneses (N = 276) que, siguiendo una reforma educativa del año 2014, diferían con respecto a la edad en la que comenzaron a recibir enseñanza en inglés. Un grupo comenzó en el primer grado (a la edad de 7-8 años) mientras que el otro grupo empezó en el tercer grado (a la edad de 9-10 años). Se utilizaron tres tests receptivos, uno de vocabulario, otro de gramática y otro que medía habilidades de discriminación fonética. Estos tests se administraron anualmente permitiendo comparar la rapidez de aprendizaje y el grado de aprendizaje de los dos grupos de niños después de uno y dos años de instrucción. Los resultados del estudio mostraron que los niños que empezaron más tarde realizaron la mayor parte de los tests mejor que los niños que empezaron más temprano. En relación a la velocidad de aprendizaje, los tests mostraron que los niños que empezaron más tarde tenían una ventaja con respecto al test gramatical mientras que el test de habilidad de discriminación fonética mostró un resultado más difuso con un parón en el desarrollo para los niños que empezaron más tarde y un mayor avance de los niños que empezaron más temprano. Finalmente, el estudio mostró que los niños alcanzaron un mayor nivel de dominio lingüístico y una mayor velocidad de aprendizaje que las niñas.

Palabras clave: Aprendizaje temprano de lenguas extranjeras; vocabulario receptivo; gramática receptiva; discriminación fonética; estudio semi-longitudinal

1. Introduction

The issue of the optimal age at which to introduce foreign languages (FL) in schools attracts the attention of educators, policy makers and second language acquisition (SLA) researchers. *The younger, the better* has been a long-standing assumption, i.e., younger children are assumed to be more efficient learners of second languages (L2), and therefore, the younger they start learning an L2, the higher levels of L2 proficiency they will attain (e.g., Hyltenstam, 1992; Krashen et al., 1979). This assumption is in line with educational policies implemented across Europe over the last decades whereby the age at which children start learning FL, predominantly English, in school has dropped (Enever, 2011; Eurydice, 2017). In Denmark, an educational law in 2014 established that children should start learning English from first grade (at age 7 to 8 years) rather than from third grade (9 to 10 years), which had been the status quo in this country since 2002.

Following the earlier introduction of modern languages in most European countries in the last decades, research on the age factor in European classroom settings has increased exponentially. Large-scale investigations have included studies across different European countries (e.g., Enever, 2011) and studies conducted in specific countries such as Spain (e.g., García Mayo & García Lecumberri, 2003 in the Basque country; Muñoz, 2006a, b in Catalonia), Switzerland (e.g., Pfenninger, 2014) and Germany (e.g., Jaekel et al., 2017).

The present study expands the range of geographical contexts that investigate the younger, the better assumption and further contributes with a unique design where we followed two groups of children who started learning English as a foreign language (EFL) within the same year (2014) in Denmark: an early starter group who began English instruction in the first grade (7-8 years old) and a late starter group who began English instruction in the third grade (9-10 years old). English proficiency was measured in three language dimensions, i.e., receptive vocabulary, receptive grammar and phonetic discrimination. Proficiency data for the two age groups were collected at the beginning of their English language instruction, after one, and after two years of instruction. This cross-sectional and short-term-longitudinal design allowed us to examine possible age group differences with respect to rate of learning and short-term proficiency (see also Fenyvesi et al., 2018). It is the first project to investigate the effect of starting age on English FL learning in Danish primary schools since Florander and Jensen (1969) and Mylov (1972).

2. Early foreign language teaching in Europe: A policy-based implementation

The teaching of FLs at younger ages in primary schools around the world has been a common feature of globalized education during the past decades (Enever, 2011). According to Copland et al. (2014), this is due to the above-mentioned assumption of the younger, the better, economic globalization, and parents' demands and expectations which put pressure on governments to introduce English earlier in schools. In Europe, the last decades have witnessed an increased attention to the linguistic and cultural diversity of the European Union (EU), resulting in a European multilingualism policy guided by the Barcelona Council of March 2002, which called for the teaching of two FLs from a very early age.

Recent large-scale studies and reports testify to the implementation of earlier FL programs in Europe. The ELLiE project (Enever, 2011) investigated the teaching of languages in primary schools in seven European countries, i.e., England, Italy, the Netherlands, Poland, Spain, Sweden and Croatia. It found a common trend towards an earlier start in FL education from the first or second year of compulsory schooling. Likewise, the Eurydice (2017) report showed that most European countries have introduced educational reforms to lower the starting age of the first compulsory FL during the last two decades. Generally, students start learning the first FL in primary school before the age of eight although in some countries, e.g., Cyprus and Poland, they even start in pre-primary education.

Given this trend in European primary school systems, a crucial question is whether the younger, the better is a correct assumption as far as instructed FL settings are concerned. This issue will be dealt with in the next section.

3. Research into the age factor in naturalistic vs. instructed foreign language settings

The influence of age on L2 learning has been, and still is, one of the most debated topics in the field of Second Language Acquisition (SLA). According to Muñoz (2008), it is important to differentiate between two distinct strands of research on this issue: one that examines the effect of age of onset (AO) in naturalistic learning settings, i.e., learning through immersion in the second language environment, and another one that investigates this issue in FL settings, i.e., in contexts where learning takes place in a formal context, typically the classroom. This differentiation is needed given the important asymmetries found in both types of contexts regarding the different amount of exposure to the target language (TL) and the different status of ultimate

attainment in the two learning contexts.[1] Whereas amount of exposure to the L2 in naturalistic contexts is in principle unlimited, exposure to the TL in the case of FL learning is generally more restricted, both in terms of the amount of input that the learner is exposed to and the number of interactional opportunities that are afforded. The notion of ultimate attainment also has a different status in the two types of learning situations. Research on age effects on naturalistic second language learning focuses on the final product of the L2 learning process and thus compares the language proficiency attained by early vs. late starters (ES vs. LS) after a minimum of 10 years of residence in the TL country (DeKeyser, 2000). In contrast, studies in FL settings compare ES vs. LS after a given number of instructional hours, which are temporally bounded by the educational system in which the study in question is embedded. Given the relatively limited amount of L2 learning that can take place in instructional settings vis-à-vis naturalistic settings, rate of learning becomes a crucial concern in FL classroom settings (Muñoz, 2008).

Research conducted in naturalistic settings has shown a rate advantage for older learners but an ultimate attainment advantage for younger learners, thus supporting the younger, the better view (e.g., Hyltenstam, 1992; Johnson & Newport, 1989; Snow & Hoefnagel-Höhle, 1978). Studies carried out in FL instructed settings has also shown a rate advantage for the LS over the ES in several language dimensions, e.g., grammaticality judgments (García Mayo, 2003; Pfenninger & Singleton, 2017), degree of foreign accent (García Lecumberri & Gallardo, 2003), production and perception in oral testing (Cenoz, 2003; Muñoz, 2014), composition writing (Cenoz, 2003) and listening and reading comprehension (Jaekel et al., 2017). However, this research has shown that unlike their naturalistic counterparts, younger learners fail to attain higher levels of L2 proficiency in the long term (Muñoz, 2011, 2014).

Two factors that can explain these results are the limited amount of L2 input that learners receive in a typical school context and the fact that the degree of contact with the FL is less intensive (Muñoz & Singleton, 2011; Singleton, 1995). Thus, younger learners would need a much longer period of time to outperform older learners. In addition, the higher cognitive maturity and metalinguistic awareness of older learners may be more adequate in instructional input-limited contexts, where younger learners cannot take advantage of their alleged advantage at implicit learning because of the scarcity of exposure to the TL (DeKeyser, 2000; Muñoz, 2006b; Nikolov & Mihaljevi Djigunović, 2006).

4. Effect of age of onset in different L2 proficiency domains

One important finding in previous research on the effects of starting age on L2 proficiency is that it does not affect all aspects of language equally. This has been well documented in studies conducted in naturalistic settings. For example, in a study with English-speaking learners of Dutch in the Netherlands, Snow and Hoefnagel-Höhle (1978) found that whereas older learner groups were especially good at syntactic and morphological acquisition and at metalinguistic ability and vocabulary, differences between the younger and older groups were less pronounced in listening comprehension and story-telling.

In studies conducted in instructed FL settings, age-related differences between the various aspects of language proficiency have also been found. For example, in a study of Swedish-speaking children starting English instruction at Grades 1, 2 and 3, Ekstrand (1982) found a positive correlation between age of onset and performance on tasks such as dictation, reading comprehension, and pronunciation but in more natural communicative tasks (i.e., free oral production), performance was strongly correlated with learners' length of residence in the host country and not with age. Cenoz (2002) found that after six years of English instruction, 11-year-old Basque-Spanish bilinguals did significantly better than 8-year-olds in tasks such as cloze tests and written composition whereas the difference between the two age groups in relation to listening comprehension was only marginal. In the BAF (Barcelona Age Factor) project in Catalonia where 4 different age groups were investigated (AO= 8, 11, 14 and +18), Muñoz (2006b) found that the largest age group differences were observed in the cloze and dictation tests, with an advantage for the late starter groups, whereas a smaller difference between the two age groups was seen in the listening comprehension task (in the initial stages of language acquisition) and in the reception measure on the oral interview (after 726 hours of English instruction).

The development of L2 phonology deserves special mention given its central role in the ongoing debate about the existence of a critical period for L2 learning. According to the Critical Period Hypothesis, this linguistic domain is among the first, if not the only one (Scovel, 1988), to be influenced by a critical period for language acquisition as it not only involves cognitive development but also neuromuscular coordination skills. Research on the acquisition of L2 phonological skills in naturalistic settings has generally supported the younger, the better view showing that even though older learners may outperform younger learners in the initial stages of FL acquisition (e.g., Snow & Hoefnagel-Höhle, 1978), younger learners often end up surpassing older learners (e.g., Krashen et al., 1982; Cook, 1991).

Research in formal instructed settings, on the other hand, has generally provided evidence for a late starting age advantage. For example, in a study examining the development of phonetic skills during six or seven years by children who had started learning English at 4, 8 and 11 years of age, García Lecumberri and Gallardo (2003) found that the older age group was rated as producing less accented and more intelligible FL speech than the two younger groups. As part of the BAF project, Fullana (2006) examined the development of phonetic skills in 4 groups of participants (AO = 8, 11, 14 and 18+ years old) by means of an auditory discrimination task and an imitation task. The results of this study showed that, especially for vowel contrasts, 8-year-olds with 200 and 416 hours of instruction obtained significantly lower correct discrimination scores than the older starting groups. This difference, however, turned out to be non-significant by the time the exposure to English instruction of all the groups had amounted to 726 hours.

The results of the above studies thus show that early starting age is not a factor that facilitates FL sound acquisition in the case of instructed FL learning. In the same vein as other dimensions of L2 proficiency, it may be the case that an early starting advantage requires much longer exposure to and contact with the TL as it is the case in naturalistic contexts (Singleton, 1995).

Age-related differences in relation to various language dimensions have been explained in terms of differences in learners' cognitive maturity. Older learners exhibit higher levels of L2 proficiency in tasks that require higher cognitive maturity such as measures of L2 syntax, morphology and other literacy-related tasks (i.e., vocabulary and reading comprehension) whereas they do not have the same advantage in tests of communicative skills that are considered to be unrelated to the cognitive/academic dimension such as oral fluency and pronunciation (e.g., Cummins, 1980; Muñoz, 2006b). Another explanation that has been offered is related to the presence of implicit vs. explicit mechanisms (Ellis, 1994) that can be at work when engaging in different language activities. Thus, Muñoz (2006b) tentatively concluded that age-related differences are greater in tasks with a higher involvement of explicit processes (e.g., cloze tests) than in those where implicit processes are arguably more present (e.g., oral comprehension task and free production tasks).

5. The study

In the present study, we compared the learning rate and short-term proficiency of two groups of Danish children – an ES group who began English instruction in first grade and a LS group who began English instruction in third grade. Children were given two standardized tests: a receptive vocabulary test (PPTV) and a receptive

grammar test (TROG) in three yearly waves of data collection and in addition a phonetic discrimination task in the second and third years. Part of the data from the first year is outlined briefly in Fenyvesi et al. (2018). Following Unsworth et al. (2014), receptive rather than production tests were chosen as primary language data given previous claims in the literature that children's production during the first two years of instruction is likely to be limited. We also designed and administered a phonetic discrimination task due to the important role that the development of phonological and phonetic skills has had in theoretical discussions and empirical research on the age factor in L2 acquisition (e.g., Flege, 2003; Piske et al., 2001).

The study addressed the following questions:

1. Will ES and LS reach different levels of FL proficiency after one and two years of instruction?

2. Will there be age and gender-related differences in the rate of FL learning during the first two years of L2 English instruction?

3. Will the learning patterns be the same for different dimensions of language proficiency, i.e., receptive vocabulary, receptive grammar, and phonetic discrimination?

5.1 Method

5.1.1. Participants

A total of 276 students (139 boys, 137 girls) participated in the study. All students began formal English instruction in 2014, so in the first year of the study, children were either in first grade (ES aged 7 to 8 years) or in the third grade (LS, aged 9 to 10 years). The children attended 6 elementary schools in the Southern region of Denmark, all located in urban settings (see Cadierno and Eskildsen, 2018 for details). Schools varied with respect to the number of instruction hours per week. In the first year of our study, the schools offered either 1 or 2 weekly English lessons. In the second year, half of the schools offered 1 weekly lesson and the other half 2 weekly lessons to the now 2nd graders, and all 4th graders received 2 weekly lessons (one lesson is 45 minutes). Given these differences, the number of instruction hours was entered as a factor in the statistical analyses.

The schools followed the guidelines on English language teaching established by the Danish Ministry of Education. The guidelines are phrased in terms of what children

should be able to do after finishing 4th, 7th, and 9th grade. By 4th grade, children should be able to participate in short and simple conversations as well as understand and write frequent words, expressions and short texts on everyday topics in English. These learning objectives apply nation-wide, but teachers are free to choose methodology. With young learners, i.e., from grade 1 through 4, the ministry's recommendations for good practice include oral and playful activities, English as the medium of instruction, and teachers' use of gestures and body language to facilitate understanding (see aus der Wieschen, 2017). In terms of educational background, the teachers typically had the required 4-year BA-degree in Education.

We know from classroom observations and recordings that the teachers sing and play with the students and use picture-books and textbooks, the latter mostly with the LS. Language choice practices differ across teachers. Most teachers use both languages to varying degrees. One case study observed a teacher who taught both ES and LS classes. He adhered to an English-only policy with the LS and used a combination of English and Danish with the ES (aus der Wieschen, 2017).

The project was reported to the Danish Data Protection Agency and to the regional ethics committee. School principals had been briefed on the project and agreed to participate. Parents were informed about the project through the schools' intranet and given the option not to let their children participate in the study.

5.1.2. Instruments

As a measure of receptive vocabulary, we used The Peabody Picture Vocabulary Test, Fourth Edition, (Dunn & Dunn, 2007), referred to here as PPVT. As a measure of receptive grammar, we used the Test for Reception of Grammar version 2, (Bishop, 2003), referred to here as TROG. Although not developed for testing EFL, the two tests have been used for that purpose (Dahl & Vulchanova, 2014; Sun et al., 2016; Unsworth et al., 2014). The two tests were administered individually and presented the original English-language items, either a word (PPVT) or a sentence (TROG). Children indicated their answers by selecting one corresponding picture out of four. Items were presented via tape recordings of an expert English speaker[2]. The PPVT has 228 items and exists in two forms, A and B, that were used in alternating years. The administration of the two tests followed the instructions in their respective manuals with the modification that the PPVT was always given starting from the first item, independently of age (as in Unsworth et al., 2014). To allow comparison of scores achieved in PPVT forms A and B, raw scores were converted to Grade Score Values (GSV) following the manual. The maximum GSV score for form A is 270 and for form B 271, both corresponding to a raw score of 228. For the TROG, the same form

with 80 items was used each year. The maximum score was 20 blocks correct (a block is a cluster of 4 items that tests the understanding of a particular linguistic phenomenon) and 80 when scored by individual items correct. We present the results for the TROG in terms of a score of total items passed (e.g., Unsworth et al., 2014). Instructions for both tests were given in Danish.

The phonetic discrimination test consisted of seven minimal pairs, all consisting of monosyllabic words: *free* vs. *three*; *sink* vs. *think*; *these* vs. *ds*; *sue* vs. *zoo*; *log* vs. *lock; ice* vs. *eyes,* and *luck* vs. *lock*. The test targets minimal pairs with sounds that are considered to be particularly difficult for Danish learners of English (Mees & Collins, 2000; see Appendix A). The test was administered in blocks of four word pairs in which the minimal pairs were given twice and the single words were repeated in a pair twice along the following lines – but note that the order in which the phenomena occur varied from minimal pair to minimal pair: 1) *free – three*; 2) *three – three*; 3) *three – free*; and 4) *free – free*. The children then had to judge whether the two items in the pairs were identical or different. A block was passed when all four pairs in a block were identified correctly. The maximum score was thus 7. For the data analysis, we derived a d-prime score which takes into account that tests where there is no difference in certain pairs may lead to response biases (Macmillan & Creelman, 1991).

5.1.3. Procedure

The children were tested individually at their school as part of a multi-day test program. As this was the only year that there was an overlap between the old and the new curricula, practical reasons dictated the start of the school year as the first data collection point. The first testing took place when the first and third-graders started English classes and the subsequent two tests exactly 1 and 2 years later.

The PPVT and TROG tests were administered at all three data collection waves whereas the phonetic discrimination task was administered in waves 2 and 3 (see Table 1). The PPVT was always administered before the TROG within the same data collection session. The phonetic discrimination task (PHON) was administered in a follow up session, typically the day after.

Table 1. Data collection points

	Fall 2014 Wave 1	Fall 2015 Wave 2	Fall 2016 Wave 3
Early Starters	Grade 1	Grade 2	Grade 3
Late Starters	Grade 3	Grade 4	Grade 5
Test	PPVT TROG	PPVT TROG PHON	PPVT TROG PHON

5.1.4. Preparation of data for analysis and presentation of model

We stacked the data for analysis with a mixed effects repeated measures model with the statistics software SAS. As a total of 276 students were included in the study and tested three times, this yielded a potential maximum of N = 3 × 276 = 828 observations. As some observations were not available for all students, the PPVT had N = 802, the TROG N = 813), the PHON N = 539), and total number of English lessons, N = 816. For the phonetic discrimination measures, the numbers of observations were lower because they were collected for the two last waves only.

The variables are summarized in Appendix B, and the specification of the model and its advantages is presented in Appendix C.

6. Results

Table 2 reports the scores at waves 1 to 3 for PPVT-GSV, TROG, and the phonetic discrimination task. We analyzed the results with the models specified in Appendix C. Table 3 shows the estimated model for the three outcome measures using a two-way fixed-effects interaction between time and starting grade. The statistical model included the children's classroom as a random effect.

Table 2. Raw scores for PPVT-GSV, TROG, and the phonetic discrimination task by starting grade and wave.

		Starting grade			
		1st grade		3rd grade	
	Wave	Mean	Standard Deviation	Mean	Standard Deviation
PPVT	1	80.82	17.07	99.77	20.38
	2	100.28	17.08	115.28	17.07
	3	109.68	23.77	131.57	23.50
TROG	1	8.78	7.69	16.69	9.00
	2	16.12	12.56	27.73	13.42
	3	26.19	15.45	44.02	17.86
PHON	2	2.92	.92	3.80	1.23
	3	3.28	1.09	3.73	1.17

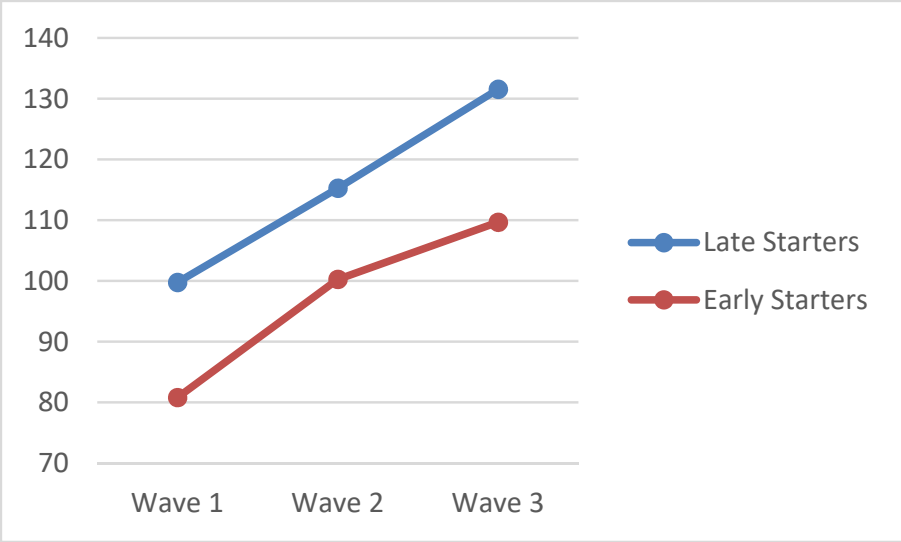
Table 3. Estimated two-way fixed-effects regression model of PPVT-GSV, TROG and the phonetic discrimination task.

Variable	**PPVT**		**TROG (items)**		**PHON (D-prime)***	
	Coefficient	*p*	Coefficient	*P*	Coefficient	*p*
Intercept	94.083	0.000	16.526	0.004	3.703	0.000
Wave2	15.220	0.000	11.202	0.000		
Wave3	32.144	0.000	27.578	0.000	-0.085	0.551
Early Starters (ES)	-17.786	0.000	-7.681	0.000	-0.888	0.000
Wave2*ES	4.123	0.109	-3.791	0.005		
Wave3*ES	-3.236	0.215	-10.200	0.000	0.457	0.013
LessonsTotal	0.035	0.352	0.001	0.979	0.001	0.742

*) Note that even though the phonetics data were collected only at waves 2 and 3, the results of the first test are here reported on the line "Early Starters."

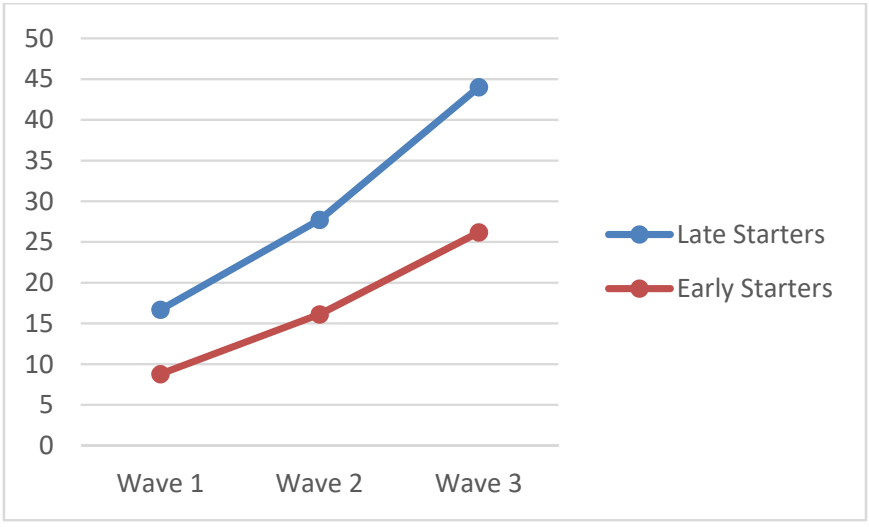
Figures 1 to 3 show the results in actual scores for ES and LS. For the PPVT, Figure 1 shows that ES began with a lower average score than LS, and that the difference between the two groups increased over time. However, while the average difference was significant ($p < 0.001$), neither the increase in difference at wave 2 (the interaction Wave2*ES, $p = 0.109$) nor wave 3 (Wave3*ES, $p = 0.215$) were significant. Thus, the difference between ES and LS remained significant at all times of test (Wave1 = 19, Wave2 = 15, Wave3 = 22 points) following the logic of this model. The ES reached the starting level of the LS at wave 2, and at wave 3, they surpassed the ES by about 10 points, the LS having a score at wave 1 of 99.77 and the ES a score at wave 3 of 109.68, $t(269) = -3.57$, $p < 0.001$.

Figure 1. PPVT-GSV scores by wave of test and starting grade



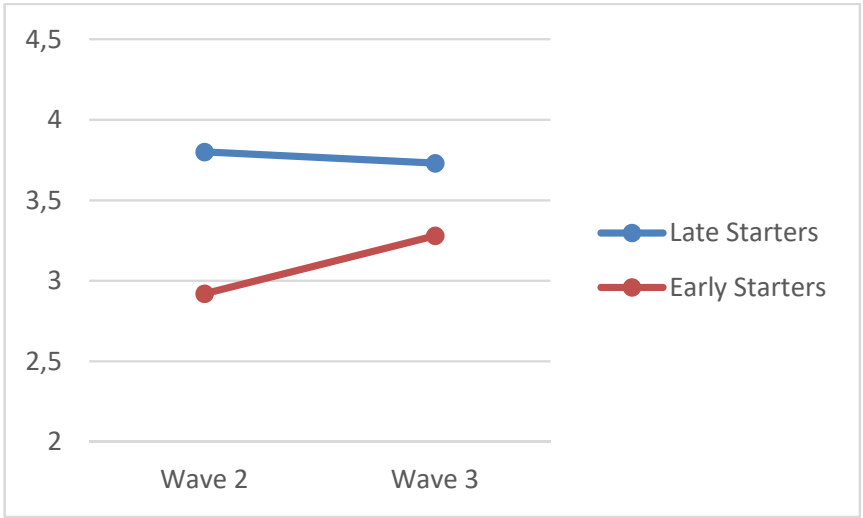
For the TROG, Figure 2 shows a similar development to the PPVT. The difference between ES and LS at wave 1 was significant ($p < 0.001$). However, as opposed to the PPVT, the increases in difference were significant at wave 2 (Wave2*ES, $p = 0.005$) and wave 3 (Wave3*ES, $p < 0.001$). Thus, the difference between ES and LS increased significantly with time (Wave1 = 8 points, Wave2 = 12 points, Wave3 = 18 points). At wave 2, the ES had reached the level of the LS at wave 1, and at wave 3 they had surpassed them by about 10 points, the LS having a score at wave 1 of 16.69 and the ES a score at wave 3 of 26.19, $t(270) = -5.83$, $p < 0.001$.

Figure 2. TROG scores, total items passed, by wave and starting grade



For the phonetics task (d-prime blocks), Figure 3 shows a significant difference between ES and LS at wave 2 (where this test was given for the first time), ($p < 0.001$). However, the difference was significantly reduced at wave 3 to about half a point (Wave3*ES, $p = 0.013$). There remained, however, a significant difference. The ES did not quite catch up with the LS, scoring 0.44 points below the LS at wave 3, the ES scoring 3.28 at wave 3 compared to the LS scoring 3.73 at wave 3, $t(256) = 3.09$, $p = 0.002$.

Figure 3. Scores on phonetic discrimination task (d-prime blocks) by wave and starting grade



Finally, as the number of lessons varied between years and classes, we included the total number of lessons of any given student across the span of the study as a predictor in the model. The average total number of lessons for the ES was 132 lessons, corresponding to 99 hours, and for the LS 166 lessons (125 hours). This difference was significant, $t(200.33) = 10.49$, $p < 0.001$. Despite the difference, the effect sizes were minimal (see Tables 2 and 3), not even reaching one tenth of a point in any of the models and not reaching significance as a predictor.

In a preliminary analysis of part of the present results, reported in Fenyvesi et al. (2018), we had found an effect of the gender of the students. Therefore, we expanded the model to include gender as a main effect. For all three outcome measures, significant differences were found between girls and boys (p varying from 0.000 to 0.030), with boys outperforming girls by an estimated 8.5 points on the PPVT-GSV and about 3.5 on the TROG. On the phonetics task, the girls did slightly better than the boys by an average of a quarter of a point. To further understand this effect, we included all interactions with gender and ran the three-way fixed-effects model shown in Table 4.

Table 4. Estimated three-way fixed-effects specifications

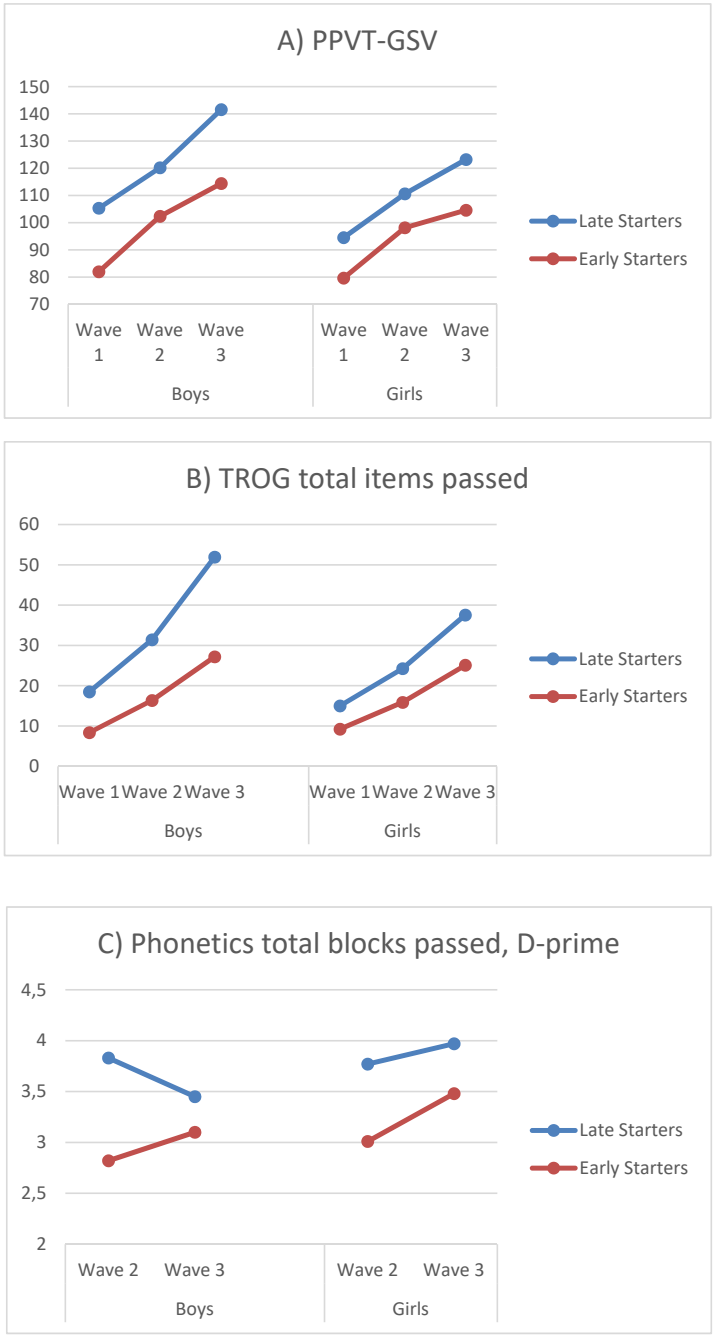
Variable	PPVT		TROG (items)		PHON (D-prime)	
	Coefficient	p	Coefficient	P	Coefficient	p
Intercept	88.148	0.000	14.806	0.009	3.737	0.000
Wave2	15.304	0.000	9.232	0.000	-	-
Wave3	13.267	0.000	13.412	0.000	0.161	0.412
Early Starters (ES)	-14.131	0.000	-5.466	0.033	-0.804	0.000
Wave2*ES	3.235	0.369	-2.437	0.194	-	-
Wave3*ES	-6.613	0.069	-4.373	0.021	0.317	0.216
Boy	10.216	0.006	3.536	0.143	-0.003	0.989
Wave2*Boy	-0.171	0.965	4.051	0.049	-	-
Wave3*Boy	8.268	0.042	6.851	0.001	-0.530	0.063
ES*boy	-7.377	0.123	-4.473	0.152	-0.177	0.509
Wave2*ES*Boy	1.745	0.733	-2.869	0.279	-	-
Wave3*ES*Boy	-2.697	0.604	-5.086	0.061	0.327	0.370
LessonsTotal	0.041	0.279	0.001	0.980	0.000	0.816

For the PPVT, see Figure 4A, boys obtained higher scores than girls by an estimated 10.22 at the outset of the study, $p = 0.006$ which did not significantly change at wave 2 ($p = .965$) but at wave 3 the boys did significantly better than the girls, by an estimated difference of 8.27 (Wave3*boy, $p = 0.042$). However, there were no significant three-way interactions.

For TROG, see Figure 4B, the development curves are quite similar to those from the two-way specification. For ES, the curves for boys and girls are practically coinciding. For LS, the development curve appears more rapid for boys than for girls by about 4 points at wave 2 (Wave2*boy, $p = 0.049$) and by about 7 points at wave 3 (Wave3*boy, $p = 0.001$), the so-called rate advantage. There was a tendency toward a statistically significant three-way interaction (Wave3*ES*boy, $p = 0.061$), which suggested that LS boys had made bigger gains than girls by wave 3. However, girls also showed a rate advantage which we confirmed by running the analysis from table 2 only for the girls and still found a significant interaction between Wave3*ES: coefficient 6.80, $p < 0.001$.

Figure 4C shows development curves for the phonetics task (d-prime blocks). There was no main effect of gender but a tendency towards a significant interaction (Wave3*boy, $p = 0.063$), providing a hint that the boys' scores drop a little less than half a point compared to the girls at wave 3.

Figure 4. Scores on PPVT (A), TROG (B), and phonetic discrimination task (C) by gender, wave, and starting grade.



7. Discussion

A change in Danish educational law provided a unique opportunity to compare the development of English receptive skills by ES and LS who had started L2 English instruction in the same year (2014). The present study examined whether the two groups of learners reached different levels of L2 English proficiency after one and two years of instruction, whether there were age and gender-related differences in the rate of FL learning during the two years of instruction, and whether the learning patterns were the same for the different dimensions of language proficiency. The development of English receptive vocabulary and grammar was tracked across three waves of data collection: wave 1 at the beginning of instruction; wave 2 after one year of instruction; and wave 3 after two years of instruction. The development of English phonetic discrimination skills was tracked at waves 2 and 3, i.e., after one and two years of instruction.

The results of the study showed that in the first wave of data collection, LS obtained significantly higher scores than ES in the two administered language tests. When tested at the beginning of their first year of English instruction, LS were better than ES in relation to receptive vocabulary and grammar. Similarly, when tested at the beginning of their second year of English instruction, LS were also superior in relation to phonetic discrimination skills. This indicates that the starting point for L2 English learning is not the same for both age groups, a finding that in the Danish context is likely due to the accumulated contact with English outside the classroom. This interpretation is supported by the results of previous studies showing that children can and do learn aspects of the L2 *before* English instruction begins (e.g., De Wilde & Eyckmans, 2017; Lefever, 2010; Kuppens, 2010).

Notwithstanding the initial difference between the two age groups, both ES and LS made significant gains in vocabulary and grammar during the first two years of English instruction. Their scores in the receptive vocabulary and grammar tests significantly increased from wave 1 to 2 and from wave 2 to 3. This finding is consistent with the findings reported by Unsworth et al. (2014) who found gains on the same proficiency tests in preschool-aged learners of EFL in the Netherlands. The pattern, however, was different for the phonetic discrimination task as only the ES (but not the LS) made significant gains between waves 2 and 3. The gain of the ES group aside, the LS still outperformed their younger peers which corroborates findings in Fullana (2006) who reported that older starters (11 and 14 years old) tended to discriminate English sounds more accurately than younger starters (8 years old) after 2.5 years of English instruction. The developmental tendency, however, is interesting and at this stage we cannot explain it, partly because we did not administer the phonetic task at wave 1 and so we do not know whether the LS had made gains in phonetic discrimination

during the first year of English instruction and then reached a plateau in relation to this particular test during the second year of instruction. It will be a question for future research to investigate whether phonetic discrimination is an area in which, in a Danish context with a high degree of extramural access to English, it will be an advantage to start early.

Did the initial advantage of LS over ES continue after one and two years of English instruction in terms of receptive vocabulary and grammar? The results of our study showed that for receptive vocabulary, the significant advantage obtained by LS over ES at wave 1 continued across waves 2 and 3. In other words, after 1 and 2 years of English instruction, children starting English classes in the 3rd grade obtained higher scores in receptive vocabulary than children starting English classes in the 1st grade. In relation to receptive grammar, the picture is slightly different. Significant interactions were found between starting grade and scores at waves 2 and 3, meaning that the initial advantage of LS over ES found at wave 1 increased over time, both after the first and second year of English instruction. This finding indicates a rate advantage for LS over ES and supports previous studies which found a rate advantage of LS over ES in grammar-related tasks (e.g., García Mayo, 2003; Jaekel et al., 2017; Florander & Jensen, 1969; Mylov, 1972; Pfenninger & Singleton, 2017). In addition, the fact that LS showed a rate advantage in relation to grammatical skills supports claims made in the literature that LS excel in tasks that require a higher degree of metalinguistic awareness and cognitive maturity, and the use of explicit learning mechanisms at which older students typically excel (DeKeyser & Larson-Hall, 2005; Muñoz, 2006b; Nikolov & Mihaljević Djigunović, 2006).

When comparing the development of the two age groups in the light of the younger, the better debate, for both receptive vocabulary and grammar, the ES group reached the starting level of the LS group after one year of instruction. That is, ES needed one year of English instruction plus potential informal learning to reach the level where LS began without prior formal instruction. In addition, after two years of English instruction, the ES surpassed the point where the LS started by about 10 points – but the LS group still showed a faster learning rate in relation to receptive grammar, which means that the gap between the two groups is widening in favor of the LS. That means that although the ES group did, in fact, do better than their later starting peers at the time of third grade, their advantage is minimal and they are still not, at least not after two years of formal English instruction, catching up with LS with respect to their performance on the PPVT and TROG tests.

Schools differed in whether they offered one or two weekly lessons during the first year of English instruction. In the second year, most schools offered 2 weekly lessons. Results showed that for vocabulary, grammar, and phonetic discrimination,

the total number of lessons over the course of the study was not a significant predictor of proficiency. Unsworth et al. (2014) tested children in a Dutch setting starting at the age of 4 1/2 years. They found that it took on average 1.5 years of instruction before a difference between 60 minutes of English classes or less per week vs 60 minutes or more per week appeared to make a difference in children's results on tests of both receptive vocabulary and grammar. However, in our study, students had participated in two years of English instruction, ranging from 68.25 hours total to 141 hours. Thus, at least in our particular context and age range, an effect of differences in instruction hours has yet to manifest itself. A partial explanation may come from classroom observations and recordings from another part of the larger project which showed that not all time allocated to English classes was necessarily spent on English (see aus der Wieschen and Eskildsen, 2019).

There was a gender gap in receptive vocabulary and grammar scores with boys generally obtaining higher scores than girls. When examining interactions between gender and rate, we noted a tendency towards an advantage in favor of boys as the two years of instruction went by. Boys both obtained higher scores and learned at a more accelerating rate than girls. This was more pronounced for the TROG where this was the case both at waves 2 and 3. A common assumption supported by some studies on young learners in the field of SLA is that females are better at learning FLs than males (e.g., Jaekel et al., 2017; Pae, 2004). Our results go against this assumption and are more in line with the results of Sylvén & Sundqvist (2012) who also found that boys outperformed girls regarding EFL vocabulary. They suggest this may be due to more time spent on gaming, a connection also made by Hannibal Jensen (2017) and Fenyvesi et al. (2018) when relating proficiency scores for a subset of our current sample to the gaming habits of the children. Research into the game-playing habits in young learners has generally shown gender differences regarding gaming frequency – with boys gaming more than girls. According to Sundqvist (2016), this may be due to gender-role stereotyping (i.e., gaming being associated with masculine culture) and lack of female characters in the games that are currently on the market. However, as noted by Carr (2005), female gaming habits may vary worldwide, and gender may thus not be a reliable predictor across the board. In fact, there are several studies conducted with younger learners that have failed to find gender differences in gaming. For example, Butler et al. (2014) found minimal gender differences in the frequency with which Japanese children aged 4 to 12 played various types of digital games. De Wilde & Eyckmans (2017) found that 11-year-old Flemish boys and girls spent a lot of time gaming (more than 1 hour per day), and as result, no gender differences were found in relation to their English proficiency.

L2 motivation is another important factor that may determine the rate and success of L2 attainment (e.g., Dörnyei & Csizér, 1998). Moreover, children's attitudes

towards the FL can be related to their gaming habits. In a recent study that examined the linguistic and attitudinal development of three Swedish children through six years of primary school, Lindgren and Enever (2017) found that the child who was the strongest in terms of English language development was the one who had more access to English at home from the start and who later sought out opportunities to engage in online gaming activities conducted in English with international partners.

Furthermore, children's attitudes and motivation can be also be gender-related. Traditionally it has been found that girls were more motivated to learn FLs than boys (e.g. Heinzmann, 2009), but Oscarson & Apelgren (2005) in a study with Swedish elementary school students found that boys were significantly more motivated than girls to learn English in order to be able to understand and use English media.

8. Conclusions, limitations and implications

This study has compared two groups of Danish school children who differed with respect to age of introduction of English instruction in school and has shown that the LS outperformed their younger peers in most tests employed. While both groups made significant gains on the PPVT and TROG, we saw a rate advantage for the LS on the TROG test. Only the phonetic discrimination test showed a diffuse picture with the LS seemingly halting in development and the ES advancing – a point which calls for further research in this specific area. The test results also generally found gender differences with boys achieving a higher level of proficiency and exhibiting a faster learning rate than girls. The phonetics test was an exception in this respect, with the girls exhibiting a higher growth rate than boys.

These results corroborate those of previous studies in not showing an advantage of introducing English earlier. However, in the specific Danish context, the younger group was the first ever to begin receiving English instruction as of the 1st grade which means that age-appropriate pedagogical tools may not yet be available to the teachers. This is different from other studies (e.g., García-Mayo, 2003; Jaekel et al., 2017; Muñoz, 2006a). At the same time, it is this historical situation that has allowed for the unique design of our study as the only one so far to study the two groups, ES vs. LS, simultaneously.

One limitation of the study is that we have only followed the children over a period of two years. Thus, we can only speak to short-term proficiency and not answer the question of whether, over time, younger learners will catch up with older learners. In future studies, we plan to follow the two groups to the end of primary school (age 16) to investigate the role of starting grade on longer-term English proficiency, testing

oral and written productive skills as well as receptive skills. Previous studies examining longer-term effects in instructed contexts, e.g., in Spain (Muñoz, 2011, 2014) have already shown a LS advantage, e.g., but it remains to be investigated whether this finding will also hold up in contexts such as Denmark where English is more readily accessible outside the classroom.

Finally, our results have pedagogical implications. The different starting points for the two groups with respect to proficiency may have consequences for what teachers are able to achieve in the classroom. As mentioned above, qualitative data point to differences in classroom practices with respect to language choice. Similarly, opportunities for interacting in English may be established at a different level with the LS, for example through a higher degree of student involvement in choosing materials as well as planning and managing activities in the classroom. Because the LS are slightly more proficient in English at the start of instruction (by around 15 to 20 points on the vocabulary and grammar tests), it may at this point in time be somewhat more straightforward for teachers in Denmark to orchestrate meaningful classroom activities with this group. Looked at from the perspective of usage-based approaches to L2 learning (Cadierno & Eskildsen, 2015), this aspect becomes all the more important because of the assumption that meaningful use is the source of learning. Age does play a role in FL teaching, it seems, but younger is not better in this context.

Acknowledgements

We thank the Danish Council for Independent Research, Humanities for funding (DFF-4001-00046). Also, we thank school principals, teachers, parents, and research assistants. Finally, and most importantly, we thank the students.

Notes

1. Even though for English, this distinction is somewhat blurred in the Danish context due to the easy access to English in everyday life, the amount of exposure and interaction in this language is presumably not as big as in second language contexts.

2. All three tests presented the test items via tape recordings. Listening to an unfamiliar voice on tape may add a cognitive load that could affect children's linguistic performance differently at different ages (e.g., Field, 2018; Winke et al., 2018). However, on the PPVT and the TROG, results showed that the ES had already at wave 2 caught up with the LS' scores from wave 1 so the

presentation method does not appear to bias the results of the ES. Further, it has been used in previous studies with children of the same age range (e.g., Unsworth et al., 2014), allowing for comparisons across studies.

9. References

aus der Wieschen, M. V. (2017). *Classroom practices in early foreign language teaching in Denmark: on the role of quantity and quality of exposure to English inside the classroom*. Unpublished Ph.D. Thesis: University of Southern Denmark.

aus der Wieschen, M. V., & Eskildsen, S. W. (2019). Embodied and occasioned learnables and teachables in early EFL classrooms. In H. T. Nguyen & T. M. (Eds.), *Conversation analytic perspectives on English language learning and teaching in global contexts: Constraints and possibilities*, Clevedon: Multilingual Matters, 31-58.

Cadierno, T., & Eskildsen, S. W. (Eds.). (2015). *Usage-based perspectives on second language learning*. Berlin / New York: Mouton de Gruyter.

Cadierno, T., & Eskildsen, S. W. (2018). The younger, the better? A usage-based approach to learning and teaching of English in Danish primary schools. *European Journal of Applied Linguistics, 6*(1), 171-182.

Basbøll, H. (2005). *The phonology of Danish*. Oxford: Oxford University Press.

Bishop, D. V. M. (2003). *Test for Reception of Grammar (version 2)*. London: Hartcourt Assessment.

Butler, T. G., Someya, Y., & Fukuhara, E. (2014). Online games for young learners' foreign language leaning. *ELT Journal, 68*(3), 265-275

Carr, D. (2005). Context, gaming pleasures and gendered preferences. *Stimulation and Gaming, 36*(4), 464-482.

Cenoz, J. (2002). Age differences in foreign language learning. *International Journal of Applied Linguistic, 135*(1), 125–142.

Cenoz, J. (2003). The influence of age on the acquisition of English: General proficiency, attitudes and code mixing. In M. del P. García Mayo & M. L. García Lecumberri (Eds.), *Age and the acquisition of English as a foreign language* (pp. 77-93). Clevedon: Multilingual Matters.

Cook, V. (1991). *Second language learning and language teaching*. London: Edward Arnold.

Copland, F, Garton, S., & Burns, A. (2014). Challenges in teaching English to young learners: Global perspectives and local realities. *TESOL Quarterly, 48*(4), 738-762.

Cummins, J. (1980). The cross-lingual dimensions of language proficiency: Implications for bilingual education and the optimal age issue. *TESOL Quarterly, 14*(2), 175-187.

Dahl, A., & Vulchanova, M. D. (2014). Naturalistic acquisition in an early language classroom. *Frontiers in Psychology, 5*, 1-9.

DeKeyser, R. (2000). The robustness of critical period effects in second language acquisition. *Studies in Second Language Acquisition, 22*, 499-533.

DeKeyser, R., & Larson-Hall, J. S. (2005). What does the critical period really mean? In J. F. Kroll & A. M. B. De Groot (Eds.), *Handbook of bilingualism: Psycholinguistic approaches* . New York: Oxford University Press. 88-108.

De Wilde, V., & Eyckmans, J. (2017). Game on! Young learners' incidental language learning of English prior to instruction. *Studies in Second Language Learning and Teaching,* 7(4), 673-694.

Dörnyei, Z., & Csizér, K. (1998). Ten commandments for motivating language learners: Results of an empirical study. *Language Teaching Research, 2*(3), 203-229.

Dunn, D. M., & Dunn, L. M (2007). *PPVT-4: Peabody Picture Vocabulary Test: Manual*. Minneapolis: Pearson Assessments.

Ekstrand, L. H. (1982). Age and length of residence as variables related to the adjustment of migrant children, with special reference to second language learning. In S. D. Krashen, R. C. Scarcella & M. H. Long (Eds.*), Child-adult differences in second language acquisition,* University of California: Newbury House Publishers. 123-158.

Ellis, N. C. (Ed.). (1994). *Implicit and explicit learning of languages*. London: Academic Press.

Enever, J. (2011). *Early language learning in Europe (ELLiE).* London: British Council.

Eurydice. (2017). *Key data on teaching languages at school in Europe*. Brussels.

Fenyvesi, K., Hansen, M., & Cadierno, T. (2018). The role of individual differences in younger vs. older primary school learners of English in Denmark. *International Review of Applied Linguistics in Language Teaching* (IRAL). https://doi.org/10.1515/iral-2017-0053

Field, J. (2018). The cognitive validity of tests of listening and speaking designed for young learners. In S. Papp & S. Rixon (Eds.), *Research and practice in assessing the English of school age young learners* (pp. 128-201). Cambridge, UK: Cambridge University Press.

Flege, J. E. (2003). Assessing constraints on second language segmental production and perception. In N. O. Schiller & A. S. Meyer (Eds.), *Phonetics and phonology in*

language comprehension and production: Differences and similarities (pp. 319-355). Berlin & New York: Mouton de Gruyter.

Florander, J., & Jensen, M. (1969). *Skoleforsøg i engelsk 1959-1965 [School experiment in English classes 1959-1965].* Stenciled: Denmark's Institute of Education.

Fullana, N. (2006). The development of English (FL) perception and production skills: Starting age and exposure effects. In C. Muñoz (Ed.), *Age and the rate of foreign language learning* (pp. 41-64). Clevedon: Multilingual Matters.

García Lecumberri, M. L., & Gallardo, F. (2003). English FL sounds in school learners of different ages. In M. del P. García Mayo & M. L. García Lecumberri (Eds.), *Age and the acquisition of English as a foreign language*. Clevedon: Multilingual Matters. 115-135.

García Mayo, M. del P. (2003). Age, length of exposure and grammaticality judgements in the acquisition of English as a foreign language. In M. del P. García Mayo & M. L. García Lecumberri (Eds.), *Age and the acquisition of English as a foreign language,* Clevedon: Multilingual Matters. 94-114

García Mayo, M. del P., & García Lecumberri, M. L. (Eds.). (2003). *Age and the acquisition of English as a Foreign Language.* Clevedon: Multilingual Matters.

Hannibal Jensen, S. (2017). Gaming as an English language learning resource among young children in Denmark. *CALICO Journal, 34*, 1–19.

Heinzmann, S. (2009). "Girls are better at language learning than boys": Do stereotypic beliefs about language learning contribute to cirls' higher motivation to learn English in primary school? *Bulletin Suisse de Linguistique Appliquee*, *89*, 19–36.

Hyltenstam, K. (1992). Non-native features of near-native speakers: On the ultimate attainment of childhood L2 learners. *Advances in Psychology, 83*, 351-368.

Jaekel, N., Schurig, M., Florian, M., & Ritter, M. (2017). From early starters to late finishers? A longitudinal study of early foreign language learning in school. *Language Learning, 67*(3), 631-664.

Johnson, J. S., & Newport, E. L. (1989). Critical period effects in second language learning: The influence of maturational state on the acquisition of English as a second language. *Cognitive Psychology, 21*(1), 60-99.

Krashen, S. D., M. A., & Scarcella, R. C. (1979). Age, rate and eventual attainment in second language acquisition. *TESOL Quarterly, 13*(4), 573-582.

Krashen, S. D., Scarcella, R. S., & Long, M. A. (Eds.) (1982). *Child-adult differences in second language acquisition.* Rowley, MA: Newbury House.

Kuppens, A. H. (2010). Incidental foreign language acquisition from media exposure. *Learning, Media and Technology, 35*(1), 65-85.

Lefever, S. (2010). English skills of young learners in Iceland: "I started talking English when I was 4 years old. It just bang... just fall into me." Paper presented at Menntakvika Conference, Reykjavik.

Lindgren, E., & Enever, J. (2017). Employing mixed methods for the construction of thick descriptions of early language learning. In J. Enever & E. Lindgren (Eds.), *Early language learning: Complexity and mixed methods* (pp. 201-221). Bristol: Multilingual Matters.

Macmillan, N. A., C. Creelman, D., & Douglas, C. (1991). *Detection theory: A user's guide*. Cambridge: Cambridge University Press.

Mees, I. M., & Collins, B. (2000). *Sound English – A pronunciation guide for speakers of Danish*. Copenhagen: Akademisk Forlag.

Muñoz, C. (Ed.). (2006a). *Age and the rate of foreign language learning*. Clevedon: Multilingual Matters.

Muñoz, C. (2006b). The effects of age on foreign language learning: The BAF project. In C. Muñoz (Ed.), *Age and the rate of foreign language learning* . Clevedon: Multilingual Matters. 1-40.

Muñoz, C. (2008). Symmetries and asymmetries of age effects in naturalistic and instructed L2 learning. *Applied Linguistics, 29*(4), 578-596.

Muñoz, C. (2011). Input and long-term effects of starting age in foreign language learning. *International Review of Applied Linguistics in Language Teaching, 49*, 113-133.

Muñoz, C., & Singleton, D. (2011). *A critical review of age-related research on L2 ultimate attainment. Language Teaching, 44*(1), 1-35.

Muñoz, C. (2014). Contrasting effects of starting age and input on the oral performance of foreign language learners. *Applied Linguistics, 35*(4), 463-482.

Mylov, P. (1972). *Skolereform i engelsk* [The reform of the English curriculum]. Copenhagen: Munksgaard.

Nikolov, M., & Mihaljević Djigunović, J. (2006). Recent research on age, second language acquisition, and early foreign language learning. *Annual Review of Applied Linguistics, 26*, 234-260.

Oscarson, M., & Apelgren, B. M. (2005). *Nationella utvärderingen av grundskolan 2003 (NU-03). Engelska. Ämnesrapport till rapport 251* [National evaluation of the primary school 2003 (NU-03). English. Curriculum report for report #251]. Stockholm: Skolverket

Pae, T.-Il. (2004). Gender effect on reading comprehension with Korean EFL learners. *System, 32*, 265–281.

Pfenninger, S. E. (2014). The misunderstood variable: Age effects as a function of type of instruction. *Studies in Second Language Learning and Teaching, 4*, 529-56.

Pfenninger, S. E., & Singleton, D. (2017). *Beyond age effects in instructional L2 learning, revisiting the age factor*. Bristol: Multilingual Matters.

Piske, T., MacKay, I. R. A., & Flege, J. E. (2001). Factors affecting degree of perceived foreign accent in an L2: A review. *Journal of Phonetics, 29*(2), 191-215.

Scovel, T. (1988). *A time to speak. A psycholinguistic inquiry into the critical period for human speech.* Rowley, MA: Newbury House.

Singleton, D. (1995). A critical look at the Critical Period Hypothesis in second language acquisition research. In D. Singleton & Z. Lengyel (Eds.), *The age factor in second language acquisition* (pp. 1-29). Clevedon: Multilingual Matters.

Snow, C. E., & Hoefnagel-Höhle, M. (1978). Age differences in the pronunciation of foreign sounds. *Language and Speech, 20*, 357-365.

Sun, H., Steinkrauss, R., Tendeiro, J., & De Bot, K. (2016). Individual differences in very young children's English acquisition in China: Internal and external factors. *Bilingualism: Language and Cognition, 19*(3), 550–566.

Sundqvist, P. (2016). Gaming and young language learners. In F. Farr & L. Murray (Eds.), *The Routledge handbook of language learning*. New York: Routledge. 446-458.

Sylvén, L. K., & Sundqvist, P. (2012). Gaming as extramural English L2 learning and L2 proficiency among young learners. *ReCALL, 24*(3), 302-321.

Unsworth, S., Persson, L., Prins, T., & de Bot, K. (2014). An investigation of factors affecting early foreign language learning in the Netherlands. *Applied Linguistics, 36*(5), 527-548.

Winke, P., Lee, S., Ahn, J. I., Choi, I., & Cui, Y. (2018). The cognitive validity of child English language tests: What young language learners and their native-speaking peers can reveal. *TESOL Quarterly, 52*(2), 274-303.

Appendix A.

The consonants targeted in the phonetic tasks are either alien in terms of the Danish phonemic inventory (/θ/ and /ð/) or (if used) not used to distinguish one word from another (/s/ vs. /z/). The phoneme /ð/ exists in Danish but it has a different place and manner of articulation (it is an alveolar approximant rather than the English dental fricative) and never used in word-initial position.

Moreover, Danish has a tendency to neutralize plosives in final position – a process known as lenition – through, e.g., deaspiration (Basbøll, 2005). So the plosive /k/ is aspirated in syllable-initial position and unaspirated in syllable-final position. This means that the final /k/ in a word such as duck is "weaker" in Danish than in English and closer to a /g/, albeit unvoiced. The test targets this potential problem in the minimal pair *log* vs. *lock*. The vowel lengthening in the English log does not have an equivalent in Danish.

Another classical problem for Danish EFL learners concerns the /ɒ/ distinction in spite of the abundance of vowels in the Danish phonemic inventory. Therefore, this distinction has also been included in the minimal pair *luck* vs. *lock*.

Appendix B

Data for the study (N=3×276)

Variable	Definition	Mean	Std	N
PPVT	PPVT score	104.10	25.02	808
TROG (items)	TROG total items correct	21.72	16.71	813
PHON (D-prime)	D-Prime score for PHON*	3.35	1.14	537
WAVE1	Indicator for Wave 1 (pre-test)	0.33	-	828
WAVE2	Joint indicator for Waves 2 and 3	0.33	-	828
WAVE3	Indicator for Wave 3 (second post-test)	0.33	-	828
ES	Indicator (1 for ES, 0 for LS)	0.60	-	828
BOY	Indicator for gender (1 for boys, 0 for girls)	0.50	-	828
LESSONS	Total lessons passed	146.04	35.22	816
CLASS	Identificator for school class	-	-	828
ID	Identificator for child	-	-	828

*) Only observed for first and second waves

Appendix C

The design to be applied is a linear regression specification, reading as

where the betas are regression coefficients, and the variables as specified in Table 1. The above specification is for PPVT; similar specifications apply to TROG scores and the phonetic task (D-prime). For the latter, specifications, given that they are only observed for waves 2 and 3, we simply omit indicators and interactions related to WAVE2_3 and thus use wave 2 as the reference.

The specification has several advantages. First, it allows for separate development curves for ES and LS by adding interaction terms. Second, it allows for inspection of significance of differences between waves and/or starting grade, which we will use and explore on in the subsequent Results section. Third, it allows for control for effects of gender and number of lessons.

We use the joint indicator WAVE2_3 rather than just an indicator for wave 2 to assess differences between wave 1 and 2 (= the coefficient for WAVE2_3) as well as the difference between wave 2 and 3 (= the coefficient for WAVE3). For simplicity, in the body of the text we refer to Wave2.

Furthermore, the specification allows for controlling unobserved effects across subgroups by using a hierarchical random effect specification. Here, we use school classes as the first level and students within school classes as the second level. Thus, the full specification becomes

$$PPVT = \beta_0 + \beta_1\ WAVE23 + \beta_2\ WAVE3 + \beta_3\ ES + \beta_4 ES * WAVE23 + \beta_5 ES * WAVE3 + \beta_6 * BOY + \beta_7\ LESSONS + e_{CLASS} + e_{PUPIL\text{-}CLASS} + e$$

By using the estimated coefficients, the average expected score for each two-way subgroup is calculated as shown in the following table:

	Early starters	**Late starters**
Wave 1	$\beta_0 + \beta_3$	β_0
Wave 2	$\beta_0 + \beta_1 + \beta_3 + \beta_4$	$\beta_0 + \beta_1$
Wave 3	$\beta_0 + \beta_1 + \beta_2 + \beta_3 + \beta_4 + \beta_5$	$\beta_0 + \beta_1 + \beta_2$

The above specification represents a two-way fixed-effects interaction between waves and ES / LS, which allows for inspecting different development curves for ES /

LS. An extension to a three-way fixed-effect interaction between waves, gender and ES / LS, which further allows for differences across gender in these development curves, is readily specified by adding interactions with gender and reads as

By using the estimated coefficients, the average expected score for each two-way subgroup is easily calculated as shown in the following table:

	Early starters - girls	**Late starters - girls**	**Early starters - boys**	**Late starters - boys**
Wave 1	$\beta_0 + \beta_3$	β_0	$\beta_0 + \beta_3 + \beta_6 + \beta_9$	$\beta_0 + \beta_6$
Wave 2	$\beta_0 + \beta_1 + \beta_3 + \beta_4$	$\beta_0 + \beta_1$	$\beta_0 + \beta_1 + \beta_3 + \beta_4 + \beta_6 \beta_7 + \beta_9 + \beta_{10}$	$\beta_0 + \beta_1 + \beta_6 + \beta_7$
Wave 3	$\beta_0 + \beta_1 + \beta_2 + \beta_3 + \beta_4 + \beta_5$	$\beta_0 + \beta_1 + \beta_2$	$\beta_0 + \beta_1 + \beta_2 + \beta_3 + \beta_4 + \beta_5 + \beta_6 + \beta_7 + \beta_8 + \beta_9 + \beta_{10}$	$\beta_0 + \beta_1 + \beta_2 + \beta_6 + \beta_7 + \beta_8$

On the translation of boundary-crossing events: Evidence from an experiment with German and Spanish translation students

Paula Cifuentes-Férez
Universidad de Murcia, Spain
paulacf@um.es

Teresa Molés-Cases
Universitat Politècnica de València, Spain
temoca1@upv.es

Abstract

This paper deals with the translation of motion events between typologically similar and different languages, a research field which has been generally approached from the Thinking-for-translating hypothesis. Here we present a student-based experiment focused on the translation of boundary-crossing events (specifically: 'manner verb + *into* + a bounded space') from English (a satellite-framed language) into German (a satellite-framed language) and Spanish (a verb-framed language). The aim is to investigate whether translation students interpret correctly and translate both the boundary-crossing and the Manner information. For this purpose, a group of German and Spanish translation students were asked to translate a series of excerpts from English narrative texts into their respective mother tongues. The results suggest that the way translation students deal with these phenomena is mainly influenced by the lexicalization patterns of their mother tongues, but the nature of the event itself and the context also seem to be key in some cases.

Keywords: Motion events, boundary-crossing, Manner, Thinking-for-translating, translation students.

Resumen

El presente artículo analiza la traducción de eventos de movimiento entre lenguas de igual y de distinto grupo tipológico, un campo de investigación que se ha estudiado principalmente en la línea de la hipótesis del Pensar para traducir. Esta contribución presenta un experimento con estudiantes en torno a la traducción de eventos de movimiento que incluyen cruce de límites (específicamente: 'verbo de manera + *into*

+ espacio delimitado') en inglés (lengua de marco satélite) y su traducción al alemán (lengua de marco satélite) y al español (lengua de marco verbal). El objetivo principal es analizar si los estudiantes de traducción interpretan correctamente y traducen tanto la información sobre el cruce de límites como sobre la Manera de movimiento. Para ello, se pidió a un grupo de estudiantes de traducción alemanes y españoles que tradujeran una serie de fragmentos de novelas en inglés a sus respectivas lenguas maternas. Si bien los resultados confirman que la forma en que los estudiantes de traducción resuelven estos fenómenos está condicionada por los patrones de lexicalización característicos de su lengua materna, la naturaleza de los eventos de movimiento y el contexto también parecen ser clave en algunos casos.

Palabras clave: eventos de movimiento, cruce de límites, Manera, Pensar para traducir, estudiantes de traducción.

1. Introduction

Languages differ in the way they filter reality through language. One of the domains which has received substantial attention in Linguistics in general and in Cognitive Linguistics in particular is motion. As argued in Filipović & Ibarretxe-Antuñano (2015: 527), among the reasons why linguists have dealt with this phenomenon, we find that motion expressions "are considered basic and omnipresent" and are "paramount to human cognition [...] and as such space and motion have been considered a crucial testing ground for linguistic behaviour".

Motion event linguistic framing and translation research has come a long way since Talmy published his seminal work on the typology of motion events (1985, 1991, 2000) and Slobin proposed his Thinking-for-speaking/translating hypotheses (1996, 2003). In the literature on the linguistic encoding of motion events (see Ibarretxe-Antuñano & Filipović, 2013, for a detailed review) and on linguistic relativity research using motion events (see, for example, Feist, 2016, for a summary), most of the existing studies compare typologically different languages, but we do find some studies which focus on languages belonging to the same typological group (e.g., Ibarretxe-Antuñano, 2004; Kopecka, 2010; Hijazo-Gascón & Ibarretxe-Antuñano, 2013; Verkerk, 2014; Lewandowski, 2018). Research on the translation of motion events pertaining to typologically similar languages is also scarce (see Cifuentes-Férez (2018) for an updated review on translation of motion events), with a few exceptions, for example, Filipović (1999, 2008) on English into Serbo-Croatian translation and vice versa, Ibarretxe-Antuñano (2003) on English into Basque and Spanish translation, Sugiyama (2005) on English into Japanese and French translation, and Lewandowski and Mateu (2016) on English into Polish and German translation. It must be noted here that the

perspective adopted in these contributions varies: while Ibarretxe-Antuñano (2003) and Sugiyama (2005) compare the translation of two target languages of the same typology (verb-framed languages), Filipović (1999, 2008) and Lewandowski and Mateu (2016) compare source and target languages belonging to the same typological group (satellite-framed). Furthermore, although research on motion encoding has extensively dealt with the boundary-crossing constraint (Aske, 1989; Slobin & Hoiting, 1994), whereby Spanish and other verb-framed languages do not allow the use of manner verbs when a boundary is crossed, that is, when an end-of-path location or a telic path is predicated. To our knowledge, research on the translation of motion has not devoted any attention to the translation of boundary-crossing events.

The present article thus aims to fill a gap in the literature on the translation of motion events by addressing the issue of the translation of boundary-crossing events and including translations from English into two languages: German (satellite-framed) and Spanish (verb-framed). In other words, the study reported in this paper focuses on the difficulties pertaining to the translation of motion events between typologically similar (English-German) and typologically different (English-Spanish) languages, with regard to the correct interpretation and rendering of boundary crossing events ('manner verb + *into* + bounded space) by student translators. The results of this study indicate that boundary-crossing events seem to be harder to interpret and translate correctly into Spanish than into German. Additionally, our data suggest that Spanish student translators might be more concerned with rendering the information conveyed in the manner verb and may overlook that a Figure has indeed crossed a boundary (expressed in prepositional phrases) and, thus, changed its location. Moreover, our findings support previous findings on the translation of Manner of motion, thus contributing to research on both inter-typological and intra-typological translation scenarios.

This paper is organized as follows. In Section 2, the context of this research will be introduced, namely, the Talmian typology of motion events (Talmy, 1991, 2000), the boundary-crossing constraint (Aske, 1989; Slobin & Hoiting, 1994), and a brief overview of research on the translation of motion events based on Slobin's (2003) Thinking-for-translating hypothesis. Then, Section 3 presents the study on the translation of boundary-crossing motion events from English into German and Spanish. After describing the research question, the hypotheses, and the methodology, we will report the results and discuss the main findings. Finally, in Section 4, some relevant conclusions and future avenues for research will be presented

2. Theoretical background

2.1. Lexicalisation patterns for motion events and the boundary-crossing constraint

It is widely known among Cognitive Linguistics scholars that Talmy (1991, 2000) distinguishes between two types of languages, depending on where the Path of motion component (the core or main component of a motion event, according to Talmy's theory) is encoded or expressed linguistically. Satellite-framed languages such as English and German typically encode Path in a satellite (usually a particle or a prefix) and Manner in the verb root. In contrast, verb-framed languages such as Spanish and French usually encode Path in the verb and Manner in adjuncts. This typological contrast is illustrated in example (1).

a. English: The bird flew out of the cage

b. German: *Der Vogel flog aus dem Käfig*

The.M bird fly.PAST.3SG from the.M.DAT cage

c. Spanish: *El pájaro salió de la jaula (volando)*

The.M bird exit.PAST.3SG from the.F cage (fly.GER)

English and German, being satellite-framed languages, express the Path component of the motion event by means of particles heading prepositional phrases (i.e., *out of the cage* in English and *aus dem Käfig* 'out of the cage' in German), while Spanish encodes the Path of motion in the verb root (i.e., *salir* 'to exit') and may express Manner in an optional constituent, as in the gerund *volando* 'flying'. However, when the motion event involves a non-telic Path, that is, when the Figure does not cross a boundary (contrary to what happens with the verbs *enter*, *cross*, or *exit*), apart from some language particularities (see, for example, Aurnague (2011), Cappelle (2012), Iacobini & Fagard (2011) and Martínez-Vázquez (2013)), verb-framed languages allow the use of manner verbs as exemplified in (2) and (3).

Paula y Alberto corrieron hasta el colegio

Paula and Alberto run.PAST.3PL up.to the.M school

'Paula and Alberto run up to the school'

Los niños bailan en el salón

The.PL kids dance.PRES.3PL in the.M dining-room

'The kids are dancing in the dining-room'

In this way, in Spanish the use of *correr* 'to run' is allowed in (2) because the preposition *hasta* 'up to' implies that the Figures (*Paula* and *Alberto*) reach the school but do not enter it, that is, they do not run *into* the school. In addition, in (3) the manner verb *bailar* 'to dance' can be used because the sentence describes an activity taking place in a specific location and no change of location is predicated.

The impossibility for Spanish to use manner verbs in boundary-crossing situations was first noticed by Aske (1989), who stated that Spanish can use a manner verb with locative path phrases as English does, but, unlike English, manner verbs cannot be employed when an end-state of the Figure is predicated. This restriction gained its name as *the boundary crossing constraint* in Slobin and Hoiting (1994), due to the authors' distinction between *path-focus* expressions, as in example (2), and *boundary-focus* expressions, as in example (1), and it is the term we are going to use throughout this paper.

Further research on motion events, such as studies carried out by Naigles et al. (1998) and Özcaliskan (2015), has suggested that the use of manner verbs is allowed in Spanish and Turkish (another verb-framed language) when the motion refers to a punctual or sudden vertical boundary crossing such as the one in *tirarse a la piscina* 'to throw oneself into the pool'. Naigles et al. (1998) also noticed that when the boundary-crossing event was horizontal, Spanish used more path verbs than when the boundary was vertical. Özcaliskan's (2015) findings indicated that Turkish speakers use manner verbs when describing quick or instantaneous acts depicted in pictures as in *leaping over a hurdle*, but not when describing motion events whose time span is longer as in *flying out of a jar*. Therefore, it seems that axis or orientation, speed of motion and/or duration of the motion event are factors which should be taken into account to better understand the boundary-crossing constraint and might be interesting to consider them when investigating both inter- and intra-typological differences in the linguistic expressions of motion events.

All these typological differences in the expression of motion events, as noted by Slobin (1991, 1996) and other scholars, have a significant impact on the narrative or rhetorical style of languages. More specifically, narratives in satellite-framed languages seem to devote more attention to how the motion event takes place, resulting in very dynamic descriptions of those events (i.e., highly rich in details of manner of motion and trajectories), whereas narratives in verb-framed languages appear to devote less

attention to the dynamics of the motion event, instead describing static settings and the result of the motion event (cf. Cifuentes-Férez, 2017).

The great interest in the linguistic expression of motion events across languages also reached research on second language acquisition. Learning a language, no matter whether it is the mother tongue (L1) or a second language (L2), requires the learner to pay attention to the aspects of experience which are easily encoded in the language being acquired. In this way, second language learners have to *rethink* for speaking in their L2 (e.g., Cadierno & Ruiz, 2006; Cadierno & Robinson, 2009; Robinson & Ellis, 2008; Cadierno, 2010) since L2 lexicalisation patterns for motion may be in competition with those of the learner's L1, especially in cases where the L1 and the L2 belong to different typological groups.

Cifuentes-Férez (2015) examined the acquisition of English motion constructions by Spanish student translators by means of inverse translation (that is, they were asked to translate a short narrative text from Spanish (L1) into English (L2)). Her data suggested that instruction on typological differences has a facilitating effect in the production of more target-like motion constructions (i.e., 'manner in the verb + path in satellite') as in the following boundary-crossing events: *Lola jumped out of the jar* and *Timón crawled out of his cage*. However, the author noticed that translators in training still face some problems with path expressions, such as *out of* (e.g. *Lola ran out [OF] the tin; Timón jumped out [OF] the cage*) and complex paths after a single motion verb, as has been also documented by other scholars (e.g., Cerda, 2010).

Inspired by a twelve picture linguistic description task used previously in Cadierno (2010) and Özcaliskan (2015), Alonso (2016) analysed the results of an interpretation task of boundary-crossing events by Spanish learners of L2 English and English native speakers. Her results showed that (a) English monolinguals chose the motion construction which is typical for satellite-framed languages; (b) Spanish learners of English use the L2 English lexicalisation pattern frequently in the scenes *dive into a pool* and *tumble into net*, which depict rapid or sudden motion; (c) pictures with *over* make participants more likely to choose manner verbs than those with *in* and *out of*; and (d) cross-linguistic influence is found especially in the case of *into/out of* + a bounded space in horizontal motion, such as *into the house*.

On the whole, research on first and second language acquisition of lexicalisation patterns of motion events has suggested that the duration of the event (punctual, rapid events vs. events extended in time) and the type of boundary crossing are important factors in the use of manner verbs in verb-framed languages, such as Spanish and Turkish. However, empirical research on the boundary-crossing constraint is still scarce and, as we shall see in the following subsection, to the best of our knowledge,

research on the translation of motion events has not yet addressed this constraint in terms of the way student translators interpret and translate boundary-crossing events from their L2 into their L1.

2.2. The Thinking-for-translating hypothesis

In his Thinking-for-translating hypothesis, Slobin (2003) discussed the consequences that the different linguistic framing of motion events may have in the translation process between similar or typologically different languages and has observed the general tendency of translators to follow the rhetorical or narrative style of the target language. In other words, translators produce target texts which sound natural to the target audience (cf. Ibarretxe-Antuñano & Filipovi, 2013; Cifuentes-Férez & Rojo, 2015; Molés-Cases, 2016, 2018; Cifuentes-Férez, 2018). His investigations of English-Spanish translation have concluded that, overall, when translating from English into Spanish, Manner information is commonly omitted and Path information is usually changed or reduced. In contrast, when translating from Spanish into English, Path is frequently kept and Manner information is either kept or even added in the target text, "apparently finding the Spanish original too bland for English readers" (Slobin, 2003: 167).

Initially, Slobin (1996, 1997) focused on the translation of English and Spanish novels, comparing originals to their published translations, and pointed to the problems that translators have to face particularly in the case of translating from English into Spanish: the greater diversity of manner verbs in English and the boundary-crossing constraint, among others. In terms of the former, when translating motion events without boundary crossing, translators have to decide whether to express the Manner information conveyed in the English novel or not. If they opt to do so, Spanish translators tend to compensate for lexical gaps in the motion verb lexicon by including adjuncts. When translating boundary-crossing events, Spanish translators accommodate the demands of their mother tongue and use path verbs in order to be able to encode the boundary-crossing, omitting or reducing the Manner information encoded in the English manner verb in the source text.

In later work, Slobin (2005) widened his scope to include another nine languages besides English and Spanish: Dutch, German, Russian and Serbo-Croatian in the satellite-framed language group and French, Portuguese, Italian, Hebrew and Turkish in the verb-framed one. He examined the translation of several fragments from Chapter 6 of the English novel *The Hobbit* by Tolkien (1937). His results further support his previous findings regarding translations from English into Spanish and vice versa, and generalise them to a larger number of languages: verb-framed languages apparently

are less concerned with Manner of motion than are satellite-framed languages, and they break complex paths or trajectories into several segments. In addition, when translating from English into verb-framed languages, it appears that there is a greater loss of both Path and Manner information than in translations from English into other satellite-framed languages.

Further studies on the translation of motion events using *The Hobbit* have been carried out by Ibarretxe-Antuñano (2003) on Basque and Spanish, Lewandowski and Mateu (2016) on Polish and German, and Alonso (2018) on Galician. Ibarretxe-Antuñano (2003) noticed intra-typological differences between Basque and Spanish both in terms of Manner and Path. With regard to the translation of manner verbs, Spanish translators tended to keep more Manner information than did Basque translators, although the translations usually involved the inclusion of a totally different Manner than the one expressed in the original English text; in other words, they tended to modulate Manner. In terms of Path, Basque translators provided richer descriptions of paths than did Spanish translators as Basque allows the use of more than two paths with motion verbs and including multiple paths is quite common. In these cases, Ibarretxe-Antuñano (2003) refers specifically to the term 'variation', rather than to 'differences'. Lewandowski and Mateu's (2016) study on translation of English into Polish and German also addressed intra-typological differences between these two satellite-languages. Their findings indicate that German exploits the satellite-framed pattern much more than Polish does, because German verb particles are more flexible than Polish prefixes when combining manner verbs and paths, and German has a richer inventory of resources for encoding Path. Moreover, the authors remarked that like Spanish, Polish uses a separate verb for each path segment. Directing our attention again towards verb-framed languages, Alonso (2018) dealt with English-into-Galician translation. Her data revealed that typological differences between English and Galician influenced the rendering of manner verbs and the choice of construction types. Accordingly, manner verbs were translated into Galician using four types of verbs, namely, manner, path, motion, and non-motion verbs and four main strategies were used when translating those manner verbs: translation of only a portion of the Manner information, translation of a different type of Manner information, translation of the same type of Manner, or omission of Manner (cf. Ibarretxe-Antuñano & Filipović, 2013; Cifuentes-Férez & Rojo, 2015). In terms of the constructions used, despite the fact that in the English source text only two constructions are found: 'manner verb + Path in satellite' and 'manner verb + subordinate clause', Galician translators resorted to a wider range of constructions, with constructions with manner verbs being the most common.

The descriptive approaches to translation between languages that are typologically similar or different have often resulted in the identification or proposal of a range of potential translation strategies and in quantification of the degree of Manner and Path information kept in the target text with regard to the source text. Ibarretxe-Antuñano

and Filipović (2013) provide an exhaustive summary of the translation strategies identified by different scholars since Slobin's (1996) ground-breaking research, which result in different outcomes ranging from a total transfer of the motion event in terms of Manner and Path to a total omission of the motion event. More recently, Molés-Cases (2016, 2018) in her investigations of the translation of Manner of motion in a German into Spanish corpus of children's and young adult literature, presents a proposal of seven translation techniques adapted to Manner, namely, lexical equivalent, paraphrase, modulation, reduction, omission, specification and addition (see Table 1), bridging the gap between translation techniques drawn from translation studies and translation strategies identified in the motion event literature commented above. Although this paper does not deal with translation techniques, a brief mention to them is included in the analysis, in an attempt to shed light on some phenomena (i.e., modulation) encountered in the data.

Table 1. Correspondence between the translation techniques for Manner proposed in the motion literature and the ones proposed by Molés-Cases (2016).

Translation techniques adapted to Manner in the motion literature	Translation techniques adapted to Manner in Molés-Cases (2016)
Total translation of Manner of motion	Lexical equivalence: Manner information is kept in the TT through lexical equivalents. Paraphrase: the TT contains an explanation about Manner in the ST.
Partial translation of Manner of motion	Reduction: Manner is partially reduced.
Omission of Manner of motion	Omission (of Manner) Omission (of motion event including Manner)
Omission of the motion event	
Specification of Manner of motion	Specification: the TT contains more specific information about Manner than the ST.
Addition of Manner of motion	Addition (of Manner) Addition (of motion event including Manner)
Translation of Manner of motion by a different sort of Manner	Modulation: Manner information in the TT is different from that in the ST.

Interestingly enough, Molés-Cases (2016, 2018) found that the Spanish translation is simpler than the original text in German in terms of Manner, and that motor

pattern seems to be most affected by translation. Furthermore, her data showed that Manner is lost in almost 30% of the corpus, but gained in 9%, which suggests that translators are using the technique of compensation (cf. Molina & Hurtado, 2002) as they could be compensating the loss of Manner in some fragments by specifying Manner details and adding Manner in other parts of the novels. On the whole, her findings are in line with previous research on the translation of motion from satellite-framed into verb-framed languages and point to some advances in the study of the translation of manner of motion events.

Despite the wealth of research on the translation of motion events, no special attention has been given to the translation of boundary-crossing events. Therefore, to our knowledge, the questions still remain unanswered as to whether student translators face any problem when understanding this type of event and whether they translate these events correctly. The following section introduces a study aimed at addressing these issues.

3. Study: research question, hypotheses and experiment

3.1. Research question and hypotheses

With this piece of research, we attempt to shed light on the translation of English boundary-crossing events taken from narrative texts into German and Spanish. The research question that this contribution aims to answer is the following: Do German and Spanish translation students encounter difficulties in the interpretation and rendering of English boundary-crossing events?

In an attempt to answer this question and drawing on previous research on the translation of motion events, in this study two sets of hypotheses are posited: the former set is related to the translation of boundary-crossing, while the latter refers to the translation of Manner of motion:

3.2. Translation of boundary-crossing:

Hypothesis 1a: In the English > German linguistic combination, German translation students are expected to interpret correctly and translate boundary-crossing in most cases. Since both English and German are satellite-framed languages, we predict that student translators will be likely to interpret boundary-crossing correctly and, thus, translate it into German.

Hypothesis 1b: In the English > Spanish linguistic combination, Spanish translation students are expected to often misinterpret boundary-crossing and do not translate it. Because in Spanish Path is typically encoded in the sentence's main verb, it is hypothesized that student translators will be more likely to focus on translating the main verbs (i.e., manner verbs) and to overlook the boundary-crossing encoded in the prepositional phrases. This hypothesis has been inspired by the continued observation of this phenomenon (loss of boundary-crossing in Spanish translations) in translation classes.

3.2. Translation of Manner:

Hypothesis 2a: In the English > German linguistic combination, German translation students are expected to translate Manner in most cases. Because both languages typically use the same lexicalization pattern (i.e., manner verb + Path in a satellite/verb particle), it is expected that German students will be likely to translate all Manner information conveyed in the boundary-crossing events.

Hypothesis 2b: In the English > Spanish linguistic combination, Spanish translation students often omit Manner. Due to the typological differences in the linguistic expression of Manner, it is expected that student translators will be likely to omit, reduce or modulate Manner information in the Spanish translation in comparison with the original in English.

4. Materials and Method

The experiment carried out consisted in requesting native German- and Spanish-speaking translation students to translate into their mother tongues five excerpts of narrative texts in English. A total of 73 translation students aged between 19 and 28 and proficiency in English between B2 and C2 participated in the experiment: 20 students from the Universität Leipzig (Germany) and 53 students from the Universidad de Murcia (Spain).[1] The students had 90 minutes to complete the task using any resource, including online dictionaries or corpora. Background information regarding the source of the texts or the research question was given to the students only after the experiment.

The materials used were five excerpts with a total of around 500 words taken from five novels originally written in English. Except for the fragment from Mockingjay, these excerpts were extracted from the COVALT corpus (Corpus Valencià de Literatura Traduïda, Universitat Jaume I), specifically from the subcorpus of original

texts in English: http://www.covalt.uji.es. The motion events included in the five excerpts (henceforth, E1, E2, E3, E4, and E5) are listed in Table 1:

Table 2. Excerpts used in the experiment.

Excerpts	Source	Author	Year
E1. *Aunt Em, badly frightened, threw open the trap door in the floor and* ***climbed down the ladder into the small, dark hole.***	*The Wonderful Wizard of Oz*	L. Frank Baum	1900
E2. *The next moment, I heard a loud shriek and a crash of something dropping, and* ***she came staggering back into the room without her light****.*	*The Ghostly rental*	Henry James	1948
E3. *Suddenly he turned tail and* ***darted into a crevice in the rock behind him****.*	*The Grizzly King*	James Oliver Curwood	1916
E4. ***We skid into a nest with a pair of soldiers****, hunching down behind the barrier.*	*Mockingjay*	Suzanne Collins	2010
E5. *The window was so wet I couldn't see, so I ran downstairs as I was and* ***slipped out the back into the garden*** *and there was the poor fellow at the end of the garden, shivering.*	*The Dead*	James Joyce	1914

As can be seen in Table 2, all motion events include Manner and boundary-crossing (encoded in the prepositional phrase '*into* + a bounded space'). We decided to focus on this type of Path because Alonso (2016) found that it was a source of major cross-linguistic influence in her research, and we considered that, as this study represents a first approximation to the translation of boundary-crossing events, it would be reasonable just to focus on one type of Path. Furthermore, it should be noticed that whereas E1 depicts downward motion, that is, motion along a vertical axis, E2, E3, E4, and E5 depict motion along a horizontal axis. In our corpus of novels, we just came across one instance of *into* + vertical axis (i.e., the one in E1); as a consequence, we are aware of the fact that the stimuli in this regard are not balanced. It is also important to highlight that in E2 the lexicalization pattern corresponds to the lexicalization pattern typical of verb-framed languages ('path verb + Manner in adjunct'), as we also wanted to explore whether this could have a facilitating effect for Spanish student translators when understanding and translating the event.

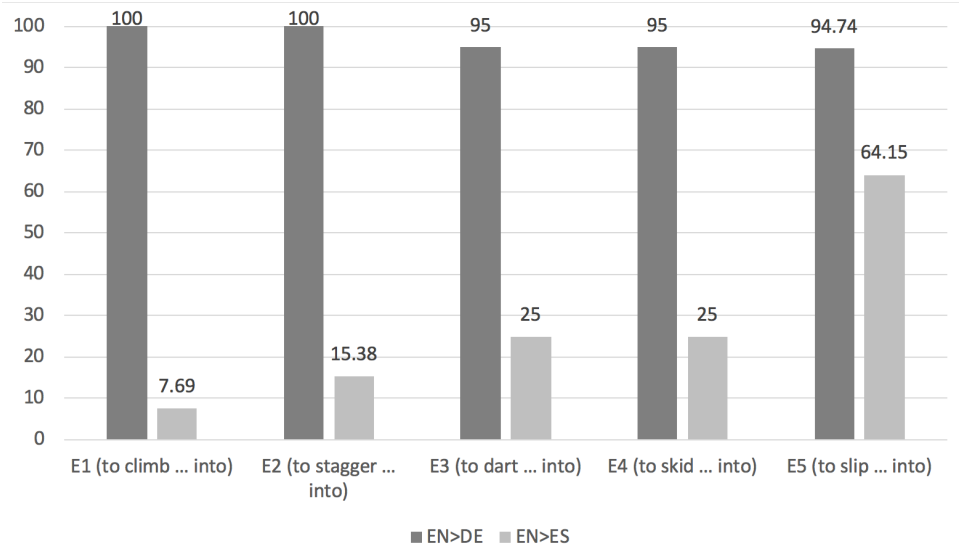
5. Results and discussion

In this section, the results of the experiment will be presented in order to test the hypotheses formulated above.

Translation of boundary crossing (hypotheses 1a and 1b)

As a reminder, as English and German belong to the same typological group and use the same characteristic lexicalisation pattern for motion, hypothesis 1a stated that German student translators would be more likely to correctly interpret and translate English boundary-crossing events. In contrast, hypothesis 1b predicted that Spanish student translators would be more prone to misinterpret boundary-crossing events and, thus, less likely to translate the crossing of boundaries in the target texts as they might be paying more attention to the meaning of the main motion verb (i.e., a manner verb in four of the five English excerpts), which in Spanish typically encodes the core component (i.e., Path) of the motion event. Figure 3 shows the results of the rendering of the boundary-crossing information for each of the five fragments and for the two linguistic combinations:

Figure 1. Results of the experiment: translation of boundary-crossing (%).



The reader should note that EN>DE stands for English>German translation and EN>ES, for English>Spanish translation. In general, as shown in Figure 1, German students seemed to have identified the boundary-crossing events correctly and they have translated them in most cases; more concretely, they have translated boundary-

crossing in all cases for E1 and E2, whereas for E3, E4, and E5 the percentage of translation of this information is over 94 % of the target texts. Therefore, in the light of these results, the research hypothesis 1a is confirmed.

E1: ([...] climbed down the ladder into the small, dark hole [...]

[...] und stieg die Leiter in das schmale, dunkle Loch hinab.

And climb.PAST.3SG the.F ladder in the.N.ACC small.N dark.N hole down

'and climbed the ladder into the narrow dark hole down'

[...] und kletterte über die Leiter *in das kleine, dunkle Loch hinunter.*

And climb.PAST.3SG over the.F ladder in the.N.ACC small.N dark.N hole down

'and climbed over the ladder into the small dark hole down'

As examples 4a) and 4b) illustrate, in the German translations of E1, it is clear that the Figure enters *into* the hole (here the German preposition *in* is used) by executing a descending motion (encoded in the verb-particles *hinab, hinunter*) involving the use of arms and legs (as expressed in the manner verbs *steigen* and *klettern*, respectively, and in the action of descending a ladder). If we have a look at other translations from other excerpts (see examples 5 and 6), it can be noticed that once again it is the preposition *in* that indicates the crossing of a boundary (*in eine Spalte, in ein Nest, in den Garten*). The fact that the English prepositions *into* and *in* (indicating crossing of boundaries) find their direct equivalents in the German preposition *in* (including the crossing of boundaries as well) could explain the high degree of translation of this type of information.

E2: [...] darted into a crevice in the rock behind him.

[...] und stürzte sich in eine Spalte im Felsen hinter ihm.

And rush.PAST.3SG himself in.ACC a.F crevice in.the.DAT rock behind him

'and rushed into a crevice in the rock behind him'

E3: We skid into a nest with a pair of soldiers [...]

Wir schlüpfen in ein Nest mit einigen Soldaten [...]

We slip.PAST.3PL in.ACC a.N nest with some.DAT soldier.PL

'we slipped into a nest with some soldiers'

On the whole, German student translators interpret and translate boundary-crossing in most of their target texts. However, the same does not seem to be the case for the translations by Spanish student translators (see Figure 1). Spanish students seemed to have failed to render the boundary-crossing events in 4 out of the 5 excerpts. Concretely, boundary-crossing was correctly translated by participants in 7.69% of cases for E1, 15.38% for E2, and 25% for E3 and E4. But, they managed to translate it in 64.15% of the cases for E5; we will deal with this particular fragment later on. These results, overall, provide support for hypothesis 1b, which stated that Spanish translators in training would be likely to misinterpret boundary-crossing events and, as result, less likely to translate the boundary-crossing information. By way of illustration, example 7 includes two translations for E1, which is the excerpt with the lowest percentage of inclusion of the boundary-crossing. These two Spanish translations describe a character who is going down or descending using a staircase until s/he reaches a small dark hole. The translations indicate that s/he is going down but not that s/he gets inside the hole, as the prepositions used by the two Spanish student translators were *hasta* 'until' and *hacia* 'towards', which do not indicate a boundary-crossing event.

E1: [...] climbed down the ladder into the small, dark hole [...]

[...] bajó *por la escalera* *hasta el pequeño* *agujero oscuro.*

Descend.PAST.3SG through the.F staircase until the.M small.M hole dark.M

'descended through the staircase until the small hole dark'

[...] *descendió* *por las escaleras* *hacia el pequeño* *agujero oscuro.*

Descend.PAST.3SG through the.PL stair.PL until the.M small.M hole dark.M

'descended through the stairs towards the small hole dark'

In addition, if we focus on example 8, which includes two translations for E4, it can be observed that, in 8a, the Spanish student translator seems to have focussed on the Manner information conveyed in the sentence's main verb (*to skid*) by using the manner verb *deslizarse* 'to slide, to skip', overlooking the fact that the characters indeed crossed a boundary (expressed by the satellite *into* in the English original text) and ended inside the nest with a pair of soldiers. In example 8b, the Spanish student translator has opted for a path verb *dirigirse* 'to direct oneself' instead of a manner verb. Because the path verb implies that the Figure reaches the Ground or reference object, but not that s/he crosses a boundary, this translation omits both the Manner information and the boundary-crossing.

E4: We skid into a nest with a pair of soldiers [...]

Entonces nos deslizamos hacia uno de los nidos con un par de soldados [...]

Then we.REFL skid.1PL towards one of the.M.PL nest.PL with a pair of soldier.PL

'Then we skid towards one of the nests with a pair of soldiers'

Nos dirigimos a un nido con un par de soldados

We.REFL direct.PAST.1PL to a nest with a pair of soldier.PL

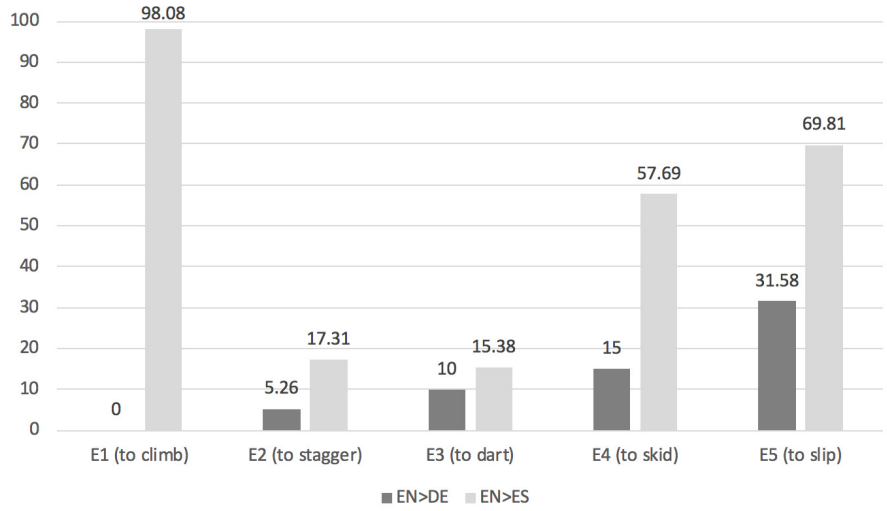
'we directed ourselves to a nest with a pair of soldiers'

In general, the examples discussed here from E1 and E4 indicate that Spanish translators in training seem to have problems with understanding and translating boundary-crossing events. It does not appear to be clear for Spanish participants that the characters ended inside a bounded space (i.e., a hole or a nest). However, to our surprise, in E5 ([...] *I ran downstairs as I was and slipped out the back into the garden* [...]), more than 60% of the Spanish participants managed to adequately translate the boundary-crossing event. The question that arises here is why E5 yielded more proper or satisfactory translations. A plausible explanation might be that the context may have played a role for the correct interpretation of the boundary-crossing event in this case, since it is very clear that the character leaves the house and goes into the garden. In spite of the fact that for E5 the boundary-crossing was translated by around 60% of Spanish participants, overall, our data for E1, E2, E3 and E4 seem to provide support for our research hypothesis 1b, which stated that in the English > Spanish linguistic combination, Spanish translation students would often misinterpret boundary-crossing and would not translate it.

Translation of Manner (hypotheses 2a and 2b)

Hypothesis 2a predicted that German student translators would be likely to translate most of the Manner information expressed in the English boundary-crossing events, whereas hypothesis 2b, on the contrary, stated that their Spanish counterparts would be more likely to omit, reduce, or modulate Manner information. Figure 2 displays the data regarding the degree of Manner lost in each of the five excerpts and in both translation scenarios (English>German and English>Spanish):

Figure 2. Results of the experiment: loss of Manner (%).



On the one hand, it can be observed that, on the whole, the amount of Manner lost in the German translations is quite low: much less than 16% for 4 out of the 5 excerpts and less than 32% for E5. Therefore, these data support research hypothesis 2a of this study, which stated that in the English > German linguistic combination, German translation students would translate Manner in most cases. The unexpected loss of Manner noticed in the translation of E5 into German deserves, however, special attention. Here, the English original fragment contained the English manner verb *to slip* but Manner is missing in around 32% of the German translations.

E5: [...] I ran downstairs as I was and slipped out the back into the garden [...]

[...] so rannte ich ohne Weiteres die Treppe runter durch die Hintertür in den Garten [...]

So run.PAST.1SG I without further.ado.N the.F stair down through the.F back.door in the.ACC garden

'so I ran just like that the stairs down through the back door into the garden'

[...] also rannte ich runter, raus und zurück in den Garten [...]

so run.PAST.1SG I down out and back in the.ACC garden

'so I ran down, out and back into the garden'

[...] also rannte ich so wie ich gekleidet war runter und in den Garten [...]

so run.PAST.1SG I as how I dress.PTCP be.PAST down and in the.ACC garden

'so I ran just like I was dressed down and into the garden'

As can be seen in the previous examples (9a, 9b, 9c), here the students have focused on the first motion event (*I ran downstairs*), in general translated by the German equivalent *rannte ich runter,* and they have omitted the second motion event (*slipped out the back into the garden),* which was the event including both boundary-crossing and Manner. This unexpected loss of Manner could be due to the following circumstances: a) the sample used is small – the translations of 20 participants were analysed –, and 31.58% of the omissions corresponds in fact to 6 cases of omission; b) this is a complex construction containing two motion events and students seem to have devoted greater attention to the first event; and c) the German lexical equivalents of *slip* (e.g., *rustchen, gleiten, schlittern, sich schleichen, schlüpfen*) are far less common in this language than the verb *rennen* 'run' (the comparison of frequencies was tested in the following corpora: COVALT (http://www.covalt.uji.es.), DWDS (https://www.dwds.de) and PAGES (https://www.corpuspages.eu)), which seems a more accessible option when translating this complex construction.

Despite the previous exception, in the German translations students tended to render most of the Manner information encoded in the English manner verbs. Examples 10 and 11 illustrate instances of the total translation of Manner in the German target texts. Here Manner is expressed in the target texts through the manner verbs *klettern* 'to climb' and *flitzten* 'to dart'. In addition, in both cases boundary-crossing is expressed by prepositional phrases introduced by the preposition *in*, and it is the context which further indicates that in example 10 motion has a vertical axis.

E1: [...] climbed down the ladder into the small, dark hole.

[...] und kletterte die Leiter in das kleine, schwarze Loch.

And climb.PAST.3SG the.F ladder in the.N.ACC small.N dark.N hole

'and climbed the ladder into the small black hole'

E3: [...] darted into a crevice in the rock behind him.

[...] und flitzte in eine Spalte im Felsen hinter ihm.

And dash.PAST.3SG in a.F crevice in.the.DAT rock behind him

'and dashed into a crevice in the rock behind him'

On the other hand, as expected due to the typological differences between English and Spanish, it can be seen that far more Manner is lost (not translated) in the Spanish target texts than in the German ones. More concretely, Manner is lost to a great extent in 3 out of the 5 excerpts, namely, E1 (98.08%), E4 (57.69%), and E5

(69.81%). This loss of Manner is even more notable in the case of E1 (*[...] climbed down the ladder into the small, dark hole*), where the Figure climbs down into a dark hole (i.e., vertical axis). This finding suggests that for this vertical boundary-crossing event in which Manner of motion is not rapid or sudden, Manner has been sacrificed in favour of Path information. As shown in example 7, the English manner verb *to climb* has been translated with the Spanish path verbs *bajar* and *descender*, both of which can be rendered into English as 'to descend'. In contrast, for E2 and E3 the degree of Manner loss is not very high (17.31% and 15.38%, respectively). As indicated earlier, E2 (*[...] she came staggering back into the room without her light*) presents the lexicalization pattern typical of verb-framed languages, which seems to have had a facilitating effect for Spanish participants leading to a higher degree of translation of Manner with respect to other fragments. The plausible reasons underlying the results of the translation of E3 (*[...] darted into a crevice in the rock behind him*) might be found in the literature on boundary-crossing motion events (Naigles et al., 1996; Özcaliskan, 2015; Alonso, 2016). This motion event includes the manner verb *to dart*, which encodes fast motion and, in this case, the motion is along a horizontal axis. Since the literature has shown that Spanish allows the use of manner verbs in fast vertical motion, this would help explain the translation of Manner here (though the motion depicted in E3 takes place along a horizontal axis). All in all, these data provide only partial support for hypothesis 2b, which predicted that in the English > Spanish linguistic combination, Spanish translation students would often omit Manner.

Focussing now on the loss of Manner information, as discussed earlier for example 8b, the Manner information encoded in *to skid* is totally omitted and the path verb *dirigirse* 'to direct oneself' is used. Similarly, in example 12, where we focussed on the second motion event encoding boundary-crossing and Manner, the English manner verb *to slip* has been translated by the Spanish path verb *salir* 'to exit' in 12a and by the deictic motion verb *ir* 'to go' in 12b, resulting in the total omission of the Manner information of this particular motion event.

E5: [...] I ran downstairs as I was and slipped out the back into the garden [...]

[...] *así que bajé corriendo por las escaleras y salí al jardín.*

So that descend.PAST.1SING run.GER by the.F.Pl stair.PL and exit. PAST.1SING to.the.M garden

'so that I descended running by the stairs and exited to the garden'

[...] *y como iba bajé corriendo las escaleras y fui hacia la parte de atrás del jardín.*

and as be.PAST.1SING descend.PAST.1SING run.GER the.F.Pl stair.PL and go.PAST.1SING towards the.F part of behind of.the garden

'so as I was I descended running the stairs and I went towards the back part of the garden'

Although this study did not aim to explore in detail the translation techniques used when rendering manner verbs, it is interesting to draw attention to the fact that Spanish participants, when they opted for including Manner in their translations, produced more instances of modulation of Manner (that is, the Manner information expressed in the target text is different from that included in the source text) than their German counterparts (18 modulations in Spanish vs. 3 instances in German; 33.96 % vs. 15%). This finding aligns well with the observations of Ibarretxe-Antuñano (2003) regarding the translation of English manner verbs by Spanish translators. Table 3 presents the cases of modulation of Manner:

Table 3. Results of the experiment: modulation of Manner.

E	Manner-verb	EN>DE	EN>ES
1	*to climb*	-	1 (*deslizarse*, 'to slip')
2	*to stagger*	1 (*sich eilen*, 'to hurry')	8 [*volver* ('return') + ADJ/ADV: *asombrada* ('shocked'), *perpleja* ('puzzled'), *pasmada* ('astonished'), *desconcertada*('disconcerted'), *atónita* ('stunned'), *impactada* ('shocked')]
3	*to dart*	-	2 (*saltar*, 'to jump')
4	*to skid* (in the sense of sliding)	-	1 (*derrapar*, 'to skid')
5	*to slip*	2 (*gehen*, 'to walk')	6 (*apresurarse*, 'to hurry'; *correr*, 'to run')
	Total	3 (15 %)	18 (33.96%)

As shown in Table 3, there has been a modulation of the manner verbs in the Spanish source texts across the five excerpts. This modulation was most frequent for E2 where the verb *to stagger* was modulated on eight occasions in the Spanish target texts by using the verb *volver* 'to return' followed by adjectives or adverbs providing information about the state of mind of the character such as *asombrada* 'shocked', *desconcertada* 'disconcerted', and so on. In contrast, as stated above, only three instances of the technique of modulation can be found in the German target texts: *to stagger* was modulated as *sich eilen* 'to hurry' in one case, and *to slip* as *gehen* 'to walk' in two instances.

6. Conclusions

The study presented in this paper has explored the issue of how German and Spanish translators in training dealt with the translation of English boundary-crossing events into their respective L1, that is, their mother tongues. This research represents a first step in the investigation of how boundary-crossing events are interpreted or understood and translated from English, a satellite-framed language, into two languages (i.e., Spanish and German) which belong to different typological groups. Our findings reveal that, unlike German translators in training, Spanish translators in training often fail to translate the boundary-crossing depicted in the excerpts and seem to face problems regarding the correct interpretation and rendering of boundary-crossing events with Manner information, suggesting that they do focus on translating the Manner in which the characters move, neglecting, as a result, the transfer of boundary-crossing information. In addition, our results suggest that Spanish translators in training are more likely to omit Manner information and to modulate Manner more often in their target texts than are German student translators.

In general, the paper provides insights into the role that the context of the excerpt can play in the correct interpretation and, therefore, the accurate rendering of the boundary-crossing event. Additionally, in line with prior research on boundary-crossing events, our data also indicate that the intrinsic nature of the events, namely, the speed as well as the vertical or horizontal orientation of the motion are relevant facets to consider when investigating this issue. Our results for the translation of the vertical boundary-crossing event showed that it was the event with the lowest rate of translation of boundary-crossing and the highest loss of Manner information in Spanish. Regarding the translation of quick or rapid motion events, it was also noticed that Spanish translators were more likely to include Manner information in their translations than with other types of motion events.

Like any piece of research, this paper would not be complete without pointing out some limitations related to the methodological choices made. As clearly stated

in Section 3, the selected excerpts only included boundary-crossing events with the preposition *into* + a bounded space, and four of the five excerpts depicted motion along a horizontal axis with just one excerpt including a vertical axis. As some differences were noted between events depicting motion along a vertical axis and those along a horizontal one, a more balanced sample of events would be desirable in future follow-ups.

Thinking about future lines of research, it would be interesting to compare our data with the results of the translation of other types of boundary-crossing, such as the ones describing coming out of a bounded space or crossing a space. Furthermore, greater attention should be paid to the difference between the vertical and horizontal axes, as well as to the duration or time span of the event so as to get a wider picture of the potential effects of all these factors on the interpreting and rendering of boundary-crossing events. Finally, it would be worthwhile to use a larger corpus of original and translated texts in order to further explore the impact of the nature of the narrated event together with the impact of linguistic framing in other verb-framed and satellite-framed languages.

On the whole, the research presented here has some pedagogical implications as it would be highly desirable for students in translation and interpreting degrees to receive formal instruction on typological differences in the lexicalization of motion events, with special attention to boundary-crossing events.

Acknowledgements

We are thankful to Michele I. Feist for her helpful feedback. We are also very grateful to the students that participated in the experiment and to our colleagues Martina Emsel, Constanze Gräsche, Elija Lutze (Universität Leipzig) and Francisco Pérez Escudero (Universidad de Murcia), for their help in carrying out the experiment at both universities.

7. References

Alonso Alonso, R. (2016). Cross-linguistic influence in the interpretation of boundary-crossing events in L2 acquisition. *Review of Cognitive Linguistics*, 14(1), 161-182.

Alonso Alonso, R. (2018). Translating motion events into typologically distinct language". *Perspectives*, 26(3), 357-376.

Aurnague, M. (2011). How motion verbs are spatial: The spatial foundations of intransitive motion verbs in French. *Lingvisticae Investigationes*, 34(1), 1-34.

Aske, J. (1989). Path predicates in English and Spanish: A closer look. In K. Hall, M. Meacham, & R. Shapiro (Eds.), *Proceedings of the fifteenth annual meeting of the Berkeley Linguistics Society* (pp. 1-14). Berkeley: Berkeley Linguistics Society.

Cadierno, T. (2010). Motion in Danish as a second language: Does the learner's L1 make a difference? In Z. Han & T. Cadierno (Eds.), *Linguistic relativity in SLA: Thinking for speaking.* (1-33) Clevedon: Multilingual Matters.

Cadierno, T., & Ruiz, L. (2006). Motion events in Spanish L2 acquisition. *Annual Review of Cognitive Linguistics*, 4, 183–216. doi: 10.1075/arcl.4.08cad

Cadierno, T., & Robinson, P. (2009). Language typology, task complexity and the development of L2 lexicalization patterns for describing motion events. *Annual Review of Cognitive Linguistics*, 7, 245–276. doi: 10.1075/arcl.7.10cad

Cappelle, B. (2012). English is less rich in Manner-of-motion verbs when translated from French. *Across Languages and Cultures*, 13(2), 173–195.

Cerda, J. P. (2010). *Estrategias utilizadas en la traducción inversa español-inglés de verbos de movimiento: un estudio en lingüística cognitiva* [Strategies in the inverse translation Spanish- English of motion verbs: a study in cognitive linguistics] (Unpublished master's thesis). Universidad de Concepción, Chile.

Cifuentes-Férez, P. (2015). Thinking-for-translating: Acquisition of English physical motion constructions by Spanish translators in training. *Cognitive Linguistic Studies*, 2(2), 302-329.

Cifuentes-Férez, P. (2017). The Impact of Typological Differences on the Perceived Degree of Dynamicity in Motion Events. *Trans*, 21, 169-185.

Cifuentes-Férez, P. (2018). Metodologías de investigación en movimiento y traducción. In A. Rojo (Ed.), *La investigación en traducción: una revisión metodológica de la disciplina*. Barcelona: Anthropos. 22-40.

Cifuentes-Férez, P., & Rojo, A. (2015). Thinking for translating: A think-aloud protocol on the translation of manner-of-motion verbs. *Target,* 27(2), 273-300.

Feist, Michele I. (2016). Minding your manners: linguistic relativity in motion. *Linguagem em (Dis)curso*, 16(3), 591-602.

Filipović, L., & Hijazo-Gascón, A. (2018). Interpreting Meaning in Police Interviews: Applied Language Typology in a Forensic Linguistics Context. *Vigo International Journal of Applied Linguistics*, 15, 67-104.

Hijazo-Gascón, A. & Ibarretxe-Antuñano, I. (2013). Las lenguas románicas y la tipología de los eventos de movimiento. *Romanische Forschungen*, 125, 467-494.

Iacobini, C. & Fagard, B. (2011). A diachronic approach to variation and change in the typology of motion event expression. A case study: From Latin to Romance. *Faits de langue-Les Cahiers*, 3, 152-171.

Ibarretxe-Antuñano, I. (2003). What Translation Tells Us about Motion: A Contrastive Study of Typologically Different Languages. *International Journal of English Studies*, 3(2), 151-166.

Ibarretxe-Antuñano, I. (2004). Motion Events in Basque Narratives. In Strömqvist, S. & L. Verhoeven (Eds.), *Relating Events in Narrative: Typological and Contextual Perspectives* (pp 89-111) New Jersey: Lawrence Erlbaum.

Ibarretxe-Antuñano, I. & Filipović. L. (2013). Lexicalisation patterns and translation. In Rojo, A. & I. Ibarretxe-Antuñano (Eds.), *Cognitive Linguistics and Translation* (pp. 251-282). Germany: Mouton de Gruyter.

Kopecka, A. (2010). Motion Events in Polish: Lexicalization Patterns and the Description of Manner. In Hasko, V., & R. Perelmutter (Eds.), *New Approaches to Slavic Verbs of Motion*. (pp 225-247) Amsterdam: John Benjamins.

Lewandowski, W. (2018). A typological approach to the encoding of motion events. In Gómez González, M.ªA., & J. Lachlan Mackenzie (Eds.), *The Construction of Discourse as Verbal Interaction*. Pragmatics and Beyond New Series, 296. (pp 45-75) Amsterdam: John Benjamins.

Lewandowski, W. & Mateu, J. (2016). Thinking for translating and intra-typological variation in satellite-framed languages. *Review of Cognitive Linguistics*, 14(1), 185-208.

Martínez-Vázquez, M. (2013). Intralinguistic variation in the expression of motion events in English and Spanish. *Lingue e Linguaggi*, 9, 143-156.

Molés-Cases, T. (2016). *La traducción de los eventos de movimiento en un corpus paralelo alemán-español de literatura infantil y juvenil*. Frankfurt am Main: Peter Lang.

Molés-Cases, T. (2018). Some advances in the study of the translation of manner of motion events. *Review of Cognitive Linguistics*, 16(1), 152-190.

Molina, L. & Hurtado Albir, A. (2002). Translation Techniques Revisited: A Dynamic and Functionalist Approach. *Meta: Translators' Journal*, 4, 498-512.

Özcaliskan, S. (2015). Ways of crossing a spatial boundary in typologically distinct languages. *Applied Psycholinguistics*, 36, 485-508.

Naigles, L., Eisenberg, A. R., Kako, E. T., Highter, M., & McGraw, N. (1998). Speaking of motion: verb use in English and Spanish. *Language and Cognitive Processes*, 13(5), 521-549.

Robinson, P., & Ellis, N. C. (2008). Conclusions: Cognitive linguistics, second language acquisition and L2 instruction – Issues for research. In P. Robinson & N. C. Ellis (Eds.), *Handbook of cognitive linguistics and second language acquisition* (pp. 489-546). New York: Routledge.

Slobin, D. I. (1996). From "thought and language" to "thinking for speaking". In Gumperz, J. & S. Levinson (Eds.), *Rethinking linguistic relativity: Studies in the social and cultural foundations of language*, vol. 1. (pp 70-96) Cambridge: Cambridge University Press.

Slobin, D. I. (1997). Mind, code and text. In J. Bybee, J. Haiman, & S. A. Thompson (Eds.), *Essays on language function and language type: Dedicated to T. Givón*. (pp 437-467) Amsterdam: John Benjamins.

Slobin, D. I. (2003). Language and Thought Online: Cognitive Consequences of Linguistic Relativity. In Gentner, D., & S. Goldin-Meadow (Eds.), *Language in Mind: Advances in the Study of Language and Thought*. (pp 157-192)Cambridge, MA: MIT Press.

Slobin, D. I., & Hoiting, N. (1994). Reference to movement in spoken and signed languages: Typological considerations. *Proceedings of the Berkeley Linguistics Society*, 20, 487-505.

Sugiyama, Y. (2005). Not all verb-framed languages are created equal: The case of Japanese. *Proceedings of the 31st Annual Meeting of Berkeley Linguistics Society*, 299–310.

Talmy, L. (1985). Lexicalization patterns: Semantic structure in lexical forms. In Shopen, T. (Ed.), *Language typology and lexical descriptions: Vol. 3. Grammatical categories and the lexicon* (pp. 36-149). Cambridge: Cambridge University Press.

Talmy, L. (1991). Path to Realization: A Typology of Event Conflation. *Proceedings of the 17th Annual Meeting of the Berkeley Linguistics Society*, 480-519.

Talmy, L. (2000). *Toward a cognitive semantics: Vol. II: Typology and process in concept structuring.* Cambridge, MA: MIT Press.

A study of L2 vocabulary acquisition under incidental and intentional conditions

Sarah Sok
University of California, Irvine
sysok@uci.edu

ZhaoHong Han
Teachers College, Columbia University
zhh2@tc.columbia.edu

Abstract

Vocabulary development is indisputably a vital aspect of second language acquisition. In spite of the abundant attention it has garnered over the past few decades, it remains unclear how adult learners fare with intentional and incidental ways of learning. The current study investigated the effects of intentional learning (via studying a word list), incidental learning (via reading), and combined intentional-and-incidental learning (via studying a word list followed by reading) conditions on 30 adult learners' second language vocabulary acquisition. Vocabulary acquisition was measured in terms of percentage gains as well as changes in the depth of vocabulary knowledge. Results showed that while both the intentional and incidental modes of learning led to vocabulary gains, the combined intentional-and-incidental condition resulted in significantly greater gains than either the intentional-only or the incidental-only condition. No significant differences were found between the incidental-only and intentional-only conditions.

Keywords: second language vocabulary acquisition, incidental learning, intentional learning, reading, word list

Resumen

El desarrollo del vocabulario es sin duda un aspecto vital de SLA. A pesar de la atención que se le ha prestado durante las últimas décadas, todavía no está claro cómo funcionan los modos intencional e incidental de aprendizaje en los adultos. Este estudio investiga los efectos de aprendizaje intencional (mediante el estudio de una lista de palabras), aprendizaje incidental (mediante la lectura) y estudio combinado

de ambos modos (estudiando una lista de palabras, seguida de una lectura) sobre la adquisición de vocabulario en la segunda lengua de 30 adultos. Medimos la adquisición en términos de porcentajes, así como en los cambios en la profundidad de conocimientos de vocabulario. Los resultados mostraron que mientras los modos de aprendizaje tanto intencional como incidental llevaron a un aumento del vocabulario, el modo combinado intencional e incidental dio como resultado un aumento significativo, mucho más alto que mediante los modos intencional o el incidental por separado. No encontramos diferencias significativas entre el modo intencional y el incidental.

Palabras clave: Adquisición de vocabulario en la segunda lengua, aprendizaje incidental, aprendizaje intencional, lectura, listas de palabras

1. Introduction

Vocabulary development is indisputably a vital aspect of second language acquisition. In spite of the abundant attention it has garnered over the past few decades, it remains unclear how adult learners fare with intentional and incidental ways of learning. The vocabulary acquisition literature shows that, to date, substantial research has been conducted on each mode of learning, and that the two modes of learning have also been compared (e.g., Barcroft 2009; File & Adams, 2010). Yet, a clear understanding of how incidental and intentional conditions contribute to L2 vocabulary learning is still lacking, resulting in areas in vocabulary teaching where research is either "not applied well" or "over-applied" (Nation, 2011: 530), and leaving L2 learners and teachers uncertain about how to overcome the burden of vocabulary acquisition (Hulstijn, 2001). Part of the confusion may be due to the ambiguity of the terms *incidental* and *intentional* and their inconsistent applications in research (Hulstijn, 2001, 2003). Another factor contributing to the confusion may be the contentious views about which approach is "better" or more effective for L2 vocabulary acquisition. For example, some L2 researchers (e.g., Krashen, 1989, 2004) maintain that incidental exposure to vocabulary during reading is the principal means of L2 lexical development. However, others contend that intentional vocabulary learning "almost always leads to greater and faster gains" (Schmitt, 2008: 341). Still others point out the futility of comparing the two approaches with the objective of determining which is superior, maintaining that both incidental and intentional learning are essential components of a well-balanced vocabulary program (Nation, 2011).

In order to lead to a deeper understanding of L2 lexical development, there is a need to disentangle these conflicting notions regarding the roles of incidental vocabulary learning (ICVL) and intentional vocabulary learning (ITVL). The current

study addressed this issue by examining L2 vocabulary acquisition under incidental, intentional, and combined intentional-and-incidental learning conditions. Vocabulary outcomes were measured with respect to both the percentage of vocabulary gains as well as the changes in depth of vocabulary knowledge.

2. Defining Incidental and Intentional Vocabulary Learning

The use of the terms *incidental* and *intentional* to describe two types of learning is widespread in L2 vocabulary research, but the terms have "often been rather loosely interpreted in common terms, not firmly rooted in a particular theory" (Hulstijn, 2003: 357). Hence, although most L2 vocabulary researchers "minimally subscribe to the meaning of the term [incidental] as referring to new knowledge or competencies that learners acquire as a 'by-product'" (Wesche & Paribakht, 1999: 176), there is persisting ambiguity with regard to the precise definitions of *incidental* and *intentional*, as evidenced by a number of studies that have attempted to deconstruct the meaning of these terms over the years (e.g., Bruton, Garcia Lopez, & Esquiliche Mesa, 2011; Hulstijn, 2001, 2003; Rieder, 2003). In one such attempt, Hulstijn (2003) delineates three ways that incidental learning has been theoretically defined in applied linguistics. The first definition of incidental learning simply denotes an absence of learner intention, and thus is "couched in negative terms" as "learning without the intent to learn" (Hulstijn, 2003: 357). The second definition refers to learning something while intending to learn something else, and differs from the first definition in that it implicates the existence of a primary learning goal that learners are intending to achieve (Hulstijn, 2003). Hulstijn explains that the origin of this definition can be traced back to the methodological traditions of early psychological experiments in which participants were told to learn a specific part of the stimuli presented, and subsequently tested (unexpectedly) on the part of the stimuli that they were *not* explicitly instructed to learn. The third definition of incidental learning is further specialized within the context of language learning and refers to "the learning of formal features through a focus of attention on semantic features" (Hulstijn, 2003: 358). Put another way, it is "learning which accrues as a by-product of language usage, without the intended purpose of learning a particular linguistic feature" (Schmitt, 2010: 29).

It should be noted that these definitions of incidental learning extend the notion of *intention* to the construct of *attention*. Particularly in the third definition mentioned above, *attention* is further specified to refer to *attention to meaning*, as opposed to *attention to form* (resonating with the distinction between meaning-focused and form-focused instruction). In the field of second language acquisition (SLA), *attending to meaning* generally refers to focusing on aspects of communication (e.g., comprehension

of the overall linguistic message) while *attending to form* refers to focusing on the formal aspects of the language, such as word forms or grammatical structures, as the object of study. However, given that most L2 researchers agree that attention is not a dichotomous construct (Gass, 1999) and that at least a small degree of attention (i.e., "noticing") is a prerequisite for any type of learning (Schmidt, 1990), a secondary distinction is often made between *focal* and *peripheral* attention in discussions of incidental and intentional learning (Ellis, 1994). Incidental learning is presumed to entail focal attention to meaning and peripheral attention to form, while intentional learning is presumed to entail the reverse: focal attention to form and peripheral attention to meaning. However, given the difficulty of quantifying attention and determining whether it is focal or peripheral, some researchers (e.g., Barcroft, 2004; Prince, 2012) claim that incidental and intentional learning should be viewed as two extremes on a continuum rather than as two distinct entities. Furthermore, they claim that no real-world linguistic task can be classified as "purely incidental" or "purely intentional" (Barcroft, 2004, p. 201). According to this view, tasks falling on one end of the incidental-intentional learning continuum would require the majority of learners' attention to be geared toward meaning and would entail that learners meet new words embedded in language-use contexts. Tasks on the opposite end of the continuum would require the majority of learners' attention to be centered on form and would entail that learners be exposed to individual words isolated from context.

In the current study, we sought to investigate the most prototypical ICVL and ITVL conditions with tasks lying on the opposite ends of the incidental-intentional continuum. ICVL was operationalized as occurring during a task in which learners primarily attended to meaning. A reading comprehension task was chosen for the current study because reading is prototypically associated with incidental learning in the field of L2 vocabulary research (Hulstijn, 2003). During the reading task, learners were not forewarned of a vocabulary test and comprehension measures were taken to ensure that learners were indeed occupied foremost with understanding the meaning of the text rather than with committing unknown words to memory. Moreover, target words were presented in context without any type of textual enhancement that would call attention to those particular word forms. As for the ITVL condition, the current study implemented a deliberate word-study task using a list, as has been done in previous studies of intentional learning (e.g., Nakata, 2008). Target words were presented along with their definitions as decontextualized, discrete items as a way of eliciting primary attention to isolated form-meaning mappings.

3. SLA Theories and Implications for Incidental and Intentional Vocabulary Learning

Researchers in the field of SLA take different positions on the roles that incidental and intentional learning play in L2 vocabulary acquisition. The disagreements stem from different theoretical perspectives about the larger process of L2 acquisition as a whole. For example, Krashen's (1987) Monitor Theory for L2 acquisition states that as learners engage in the process of comprehending meaning, language acquisition occurs via the interaction between the linguistic information encoding communicative meaning and learners' innate language acquisition faculty. More specifically, the theory posits that ample exposure to *comprehensible input (i+1),* or input that is just one step above the learner's current level of competence, is an essential and sufficient environmental ingredient for language acquisition. With regard to vocabulary, Krashen (1987, 1989) asserts that L2 learners could virtually acquire *all* the vocabulary they need through encountering comprehensible input during extensive reading. During the early years of L2 vocabulary research, this view cast an especially favorable spotlight on ICVL as opposed to ITVL (Krashen, 1989, 2004, 2013), leading to "the seductive prospect that, provided learners [have] access to sufficient comprehensible input, L2 vocabulary acquisition would largely take care of itself, without the need for any substantial pedagogical intervention" (Read, 2004, p. 147).

On the other hand, Skill Acquisition Theory (Anderson, 1982) and its application to SLA (see, e.g., DeKeyser, 2015) suggest that language acquisition originates in conscious processes that lead to the development of declarative knowledge, which, through practice, gives rise to procedural and ultimately, automatized knowledge. In other words, the type of knowledge gained as learners engage in deliberate, conscious, effortful learning, utilizing memorization and rehearsal techniques, is presumed to give rise to explicit knowledge of form-meaning links and, eventually, to fluent access and appropriate use of these form-meaning links. In contrast to Krashen's Monitor Theory (1987), this view strongly espouses ITVL as opposed to ICVL.

A third view comes from usage-based (i.e., connectionist, emergentist) theoretical approaches (Ellis & Wulff, 2015), which has been gaining increasing traction in SLA. A usage-based view sees language as a complex adaptive system emerging over time as learners' perceptual system and simple learning mechanisms interact with linguistic input. Accordingly, language learning is deemed a primarily implicit process, made possible by the interaction between learners' general cognitive mechanisms and exposure to language input during usage. This view holds that plentiful opportunities for meaning-focused input processing should be made available, while acknowledging that "form-focused instruction can help…by recruiting learners' explicit, conscious

processing to allow them to consolidate unitized form-function bindings of novel L2 constructions" which allows for "subsequent implicit processing [to] update the statistical tallying of its frequency of usage and probabilities of form-function mapping" (Ellis, 2013: 376). In other words, explicit knowledge can be helpful in that it can aid subsequent implicit learning of linguistic constructions, acting as a "support" or "scaffolding device." Applied to vocabulary learning, this perspective encourages a combination of both ITVL and ICVL. Explicit knowledge of new words gained through ITVL can provide a scaffold for subsequent implicit processing of these words when they are encountered during ICVL conditions. Furthermore, during ICVL conditions in which learners are exposed to words in context, learners' perceptual systems may take into account a variety of information beyond words' form-meaning links, such as information about how words are placed in semantic and syntactic relationships to other words.

4. Main Findings on Incidental and Intentional L2 Vocabulary Acquisition

A number of empirical studies conducted on ICVL have found that vocabulary gains from reading a text are modest at best (Day, Omura & Hiramatsu, 1991; Dupuy & Krashen, 1993; Horst, Cobb & Meara, 1998; Pitts, White & Krashen, 1989, Waring & Takaki, 2003). Studies reported average gains of only one to five words learned per text read (Hulstijn, 1992; Luppescu & Day, 1993; Knight, 1994), despite the fact that the texts read exceeded 1,000 words (Hill & Laufer, 2003). A review of previous studies by Waring and Nation (2004) found that on average only about one tenth of the target words tested were learned. Similarly, Waring and Takaki (2003) showed that words encountered more than 18 times still only carried a 10-15% chance of being retained after three months. Taken together, these findings confirmed Hulstijn's (1992: 122) assessment that "the retention of word meaning in a true incidental learning task is very low indeed".

In contrast, a number of early psychological experiments (e.g., Eagle & Leiter, 1964; Postman & Adams, 1956; Thorndike, 1908; Webb, 1962) as well as more recent research in L2 vocabulary acquisition (e.g., Laufer, 2005; Schmitt, 2010, Won, 2008) have shown that ITVL is effective and efficient. For example, Barcroft and Rott (2010), who investigated the rate of vocabulary learning from studying bilingual word lists with 24 target words, found as much as 41.9% gains for L2 learners of German and 43.8% gains for L2 learners of Spanish. Likewise, Nakata (2017) reported that Japanese-L1 EFL learners, who participated in paired associate learning of target words with either one, three, five, and seven retrievals, were able to gain between 59.1% and 93% of target words on translation tests. Although these gains decreased over time, scores on

the delayed posttests showed that substantial gains were retained even after four weeks had elapsed. For example, the group with one retrieval still showed average scores of 24.2% and 59.6% on the L1-to-L2 and L2-to-L1 translation tests, respectively, and the group with seven retrievals showed average scores of 48.8% and 83.3% on the two translation tests.

Some studies have compared the effects of ICVL and ITVL and have shown that ITVL is more effective (e.g., Ahmad, 2011; Barcroft, 2009; File & Adams, 2010; Qian, 1996). By way of illustration, Qian (1996) compared the effects of a "contextualized vocabulary" condition (ICVL) and a "decontextualized vocabulary" condition (ITVL) on Chinese-L1 EFL students' vocabulary acquisition. Participants in the contextualized vocabulary group read a passage in which target words were italicized and also listed below the text. After reading the text, participants worked out the meaning of the italicized words from the context and answered comprehension questions about the passage. Participants in the decontextualized vocabulary group received explicit instruction on the target words. The instruction involved the teacher writing each target word on a blackboard with a definition or synonym, students reading aloud the words and their definitions in unison a total of three times, and students being asked to recall the target words and definitions. The decontextualized group consistently outperformed the contextualized group on immediate, one-week delayed, and three-weeks delayed posttests. Percentages of target word retention for the decontextualized group were 75% for the first posttest, 67% for the second posttest, and 61% for the third posttest; for the contextualized group, retention was 66%, 58%, and 51% for the first, second, and final posttests, respectively.

Studies (e.g., Ertürk, 2016; Hulstijn, Hollander, & Greidanus, 1996; Knight, 1994; Luppescu & Day, 1993; Paribakht & Wesche, 1997; Sonbul & Schmitt, 2010; Zimmerman, 1997) that have compared "reading-only" with "reading-plus conditions" seem to also support the notion that ITVL is more effective than ICVL. For example, Luppescu and Day (1993) and Knight (1994) both compared reading-only conditions with reading-plus-dictionary conditions during which participants were able to look up unknown vocabulary items in a dictionary while they read. Both studies showed significantly greater vocabulary gains for the reading-plus-dictionary conditions. In a more recent study, Ertürk (2016), comparing a reading-only condition with a reading-plus-vocabulary-study condition found that the reading-plus condition yielded greater vocabulary gains. Speaking in terms of the incidental-intentional learning continuum, the reading-plus-dictionary conditions in Luppescu and Day's (1993) and Knight's (1994) studies arguably qualify as "less incidental" than the reading-only conditions because the act of looking up words can cause more attention to be paid to the formal features of words. The reading-plus-vocabulary-study condition in Ertürk's (2016) study

illustrates a hybrid of incidental learning and intentional learning. Together, these studies seem to underscore that "less incidental" and "incidental-plus-intentional" conditions tend to be more conducive to L2 lexical development than incidental-only conditions.

Still, many SLA researchers continue to hold out that incidental learning will ultimately prove uniquely beneficial for L2 vocabulary development, contributing, in particular, to the depth of vocabulary knowledge, in addition to consolidation and enhancement of this knowledge. While most studies using traditional vocabulary outcome measures (e.g., meaning-translation or multiple-choice tests) were insensitive to these gains, studies that used more fine-grained measures of vocabulary acquisition were able to show these incremental gains in word knowledge (e.g., Daskalovska, 2011; Pellicer-Sánchez, 2015; Pellicer-Sánchez & Schmitt, 2010; Pigada & Schmitt, 2006; Webb, 2007). For example, Webb's (2007) study of incidental vocabulary acquisition through reading employed a vocabulary test that measured receptive and productive knowledge of five different word knowledge dimensions: orthography, association, grammatical functions, syntax, and form-meaning link. The results indicated that knowledge in at least one of the five different dimensions increased with each increase in encounter with a word in the context of reading.

Further, despite the widely acknowledged effectiveness of ITVL, it has been noted that lexical gains from intentional learning can differ drastically based on several variables. Nakata (2008), for instance, demonstrated that the nature of the ITVL task may make a difference in the amount of vocabulary gains that are retained. In his study, Japanese-L1 high school EFL students deliberately learned words in one of three conditions: (1) studying a word list, (2) studying word cards, and (3) studying words through an interactive computer program. No significant differences were found between the three groups' immediate posttest scores (gains were 82% for the word list group, 83.1% for the word cards group, and 84.3% for the computer group), but during the four-weeks delayed posttest, the computer group performed significantly better than the word list group (gains were 26.1% for the word list group, 32% for the word cards group, and 38.2% for the computer group), implying that interactive computer-mediated learning may be more effective for vocabulary retention than the relatively more traditional method of studying word lists. Research has also shown that certain strategies such as spaced retrieval, mnemonic or other rehearsal techniques, reordering of target word items, use of the L1, and utilization of pictures can significantly enhance the effectiveness of deliberate vocabulary learning (Nation, 2011).

5. The Present Study

Building on emerging insights on the complementarity of ITVL and ICVL in promoting L2 vocabulary development, the present study explored three questions:

1. What are the effects of intentional-only (IT), incidental-only (IC), and combined intentional-and-incidental (IT-IC) learning conditions on L2 learners' vocabulary gains?

2. What differences in vocabulary gains exist between the IT, IC and IT-IC conditions?

3. What changes in depth of vocabulary knowledge occur as a result of the IT, IC, and IT-IC conditions?

A within-subject pretest-posttest design, in which all participants were exposed to the three experimental conditions, was implemented. The three experimental conditions included: intentional-only (IT), incidental-only (IC), and combined intentional-and-incidental (IT-IC). The IT condition involved a 25-minute word-study task in which the participants encountered the words in a vocabulary sheet that listed the words with their definitions. The IC condition involved a reading task segmented and distributed over three separate days, in which the participants met the target words in context during reading. The IT-IC condition involved both a 25-minute word-study task followed by a reading task distributed over three separate days.

5.1. Participants

Participants were 30 adult English as a Second Language (ESL) learners who were attending a community language school in New York City at the time of the study. They came from diverse first language backgrounds and their mean age was 31.53 (SD = 9.96) years, with the youngest participant being 20 and the oldest 67 years old. The participants' L2 proficiency level was beginning (n=6) and intermediate (n=24), as determined by an in-house placement test administered at the community language school. The participants were recruited for the current study through advertisements at the language school. They were initially told that the study entailed reading activities and English language worksheets; they were not told that the focus of the study was on vocabulary acquisition until the end of the research study.

5.2. Instruments

5.2.1. Reading Text

A graded reader, *The Elephant Man* (Vicary, 1989), which is classified as Level 1 in the Oxford Bookworms Library, was used during the reading tasks. The original version of the text was slightly modified through the insertion of target words throughout the text. The modified text contained 5,429 words, 594 word types, and 414 word families. Knowledge of the 2,000 most frequent word families was shown to account for 94.05% lexical coverage, and knowledge of the 4,000 most frequent word families provided 95.08% lexical coverage. The text was assumed to be at an appropriate level for the participants in the current study based on surveys of their class textbooks and consultations with their English language instructors prior to the study. Moreover, during the study, participants scored well on the reading comprehension questions (mean of 91% correct). In a post-study questionnaire, all participants indicated that they understood over 75% of the text they read. Moreover, 77% of the participants (n=23) marked the reading text as "easy," or "very easy," while 23% of the participants (n=7) marked it as "in the middle;" none of the participants indicated that the reading text was difficult. All participants indicated that they had enjoyed the story.

In order to confirm that reading the text represented an incidental learning condition, participants were also asked in the post-study questionnaire whether they had known that the purpose of the study was to acquire vocabulary words. Participants unanimously reported that they had not known that the study was about vocabulary acquisition, and that they had not focused on learning vocabulary unless explicitly instructed to do so (e.g., in the intentional learning condition).

5.2.2. Vocabulary List

The material used for the word study tasks were two printed handouts, one containing a list of the target words in the IT condition and one the target words in the IT-IC condition. Each target word was presented along with information on the grammatical class (e.g., noun, verb, or adjective) and the most commonly used definition of the word. Prior to the study, all definitions were reviewed by the participants' English language instructors to ascertain that they only contained words that the participants were likely to know.

In order to confirm that studying the vocabulary sheets represented an intentional learning condition, participants were asked in the post-study questionnaire whether they had attempted to commit the words to memory during the word study task. All

participants indicated that they had deliberately attempted to learn the words that were presented on the vocabulary sheets (e.g., through memorization or rehearsal techniques and strategies).

5.2.3. Target Words

There were 30 target words in the present study, 10 per experimental condition. In order to select the target words, content words that appeared frequently throughout the reading text were identified and compiled into an initial list of 50 words. (As the text was a graded reader, a large number of content words were used repeatedly in the text). Based on this initial list of words, a second list of words containing a synonym for each of the 50 original words was created. This new list of synonyms was made up of words that beginning and intermediate ESL learners were unlikely to know, based on consultations with experienced teachers at the community language school where the participants were enrolled. After further consultation with the teachers, the list of synonyms was pared down to 30 words comprising 1 adverb, 14 adjectives, 6 nouns, and 9 verbs. These 30 words were randomly assigned to one of the three experimental conditions. The purpose of the random assignment was to ensure that any differences between the sets of target words assigned to each condition were not systematic. The authors considered assigning words to each condition based on word properties (e.g., length, grammatical class, semantic complexity, frequency category in word lists), but ultimately the method of random assignment was selected because only preliminary research (e.g., Kweon & Kim, 2008) is available on the extent to which word characteristics may affect the learnability of lexical items.

The 10 target words assigned to the IC condition and the 10 target words assigned to the IT-IC condition were inserted in the reading text (in Chapters 2 through 7) in the place of their respective synonymous words. Because frequency of encounters with target words has been found to be a significant variable that affects learning (Folse, 2006; Webb, 2007), the frequency with which the target words appeared in the text was controlled so that half of the IC and IT-IC target words appeared in 3 different locations in the text and the other half in 8 different locations.

The target words in the IT condition included: *scrutinize, solicit, chortle, traipse, sluggish, gruesome, congenial, alcove, knapsack,* and *demeanor.*

The target words in the IC condition included: *oblige, hastily, lingering, indignant, apprehensive, content, dwell, exquisite, tranquil,* and *simulacrum.*

The target words in the ITIC condition included: *deceased*, *cherish*, *diminutive*, *surmise*, *gargantuan*, *grin*, *sympathetic*, *infirm*, *abode*, and *sanatorium*.

5.3. Vocabulary Outcome Measure and Scoring

The current study employed Paribakht and Wesche's (1997) Vocabulary Knowledge Scale (VKS) as pretest and posttest to assess the participants' vocabulary acquisition. The VKS is a test that was originally created to gauge lexical development, the underlying assumption being that lexical knowledge grows incrementally from no knowledge of a word to partial mastery to full mastery. Each test item in the VKS is presented with the following five categories:

I. I don't remember having seen the word before.

II. I have seen the word before, but I don't know what it means.

III. I have seen this word before, and I think it means _____. (synonym or translation)

IV. I know this word. It means _____. (synonym or translation)

V. I can use this word in a sentence: _____. (Write a sentence.) (If you do this section, please also do Section IV.)

(Paribakht & Wesche, 1997)

Paribakht and Wesche (1997) laid out the following scoring scheme for the VKS, which prescribes points on the basis of self-report (Categories I and II) as well as performance data (Categories III, IV, and V). One point is awarded if the test-taker marks Category I, two points are awarded for Category II, and three points are awarded if a word's meaning is correctly identified through an acceptable English synonym or definition in Category III or IV. In the case that a test-taker attempts to fill out Category III or IV but fails to provide an acceptable synonym or definition, only two points are awarded. Four points are awarded if the word's meaning is correctly identified in Category III or IV *and* the word is used in a semantically appropriate way in a sentence in Category V. Finally, five points are awarded if the word's meaning is correctly identified in Category III or IV *and* the word is used in both a semantically and grammatically appropriate way in a sentence in Category V.

Due to its scalar format and scoring scheme, the VKS has the benefit of distinguishing between different abilities based on the test-taker's depth of vocabulary knowledge, for example, from being able to recognize the form of a word (Category II) to being able to recall its form-meaning connection (Categories III and IV) to being able to use the word in a sentence context (Category V). The VKS is, therefore, more sensitive to changes in the depth of vocabulary knowledge than traditional meaning-recall tests or meaning-recognition tests. Moreover, the test-retest reliability of the VKS was found to be high (0.89 on content words and 0.82 on discourse connectives) (Wesche & Paribakht, 1996) and the test was shown to be useful in gauging the beginning stages of vocabulary learning and retention (Paribakht, 2005; Wesche & Paribakht, 2009).

Over the years, the VKS has become a widely used measure in L2 vocabulary research, but it has also met with several criticisms (Read, 2000; Schmitt, 2010). One of the criticisms levelled at the VKS is related to its inability to probe the complexity of vocabulary knowledge, which comprises multiple dimensions of form, meaning, and function; it has been pointed out that multiple scales rather than a single scale may be needed to adequately represent this construct (Read, 2000). Moreover, researchers (e.g., Meara, 1990) have noted that at times the ability to use a word in a sentence context can precede the ability to produce a definition or synonym of the word, contrary to the assumption inherent in the VKS that the former always precedes the latter. This criticism also extends to the VKS scoring method. Researchers have pointed out that ordinal values 1 through 5 should not be assigned to categories that can very well be nominal, contesting the view that vocabulary knowledge always develops in a linear fashion across the distinct stages outlined in the VKS.

In order to take advantage of the test's ability to differentiate between levels of vocabulary knowledge and at the same time avoid depending on unverified assumptions about the progression of vocabulary knowledge, the current study scored the VKS pretest and posttest using two methods. First, the VKS pretest and posttest were scored using the traditional dichotomous scoring method based on known and unknown words. If a word's meaning was correctly identified, one point was awarded, but if a word was not correctly identified, zero points were awarded. This scoring method yielded raw gains in target vocabulary. Second, the VKS pretest and posttest were also scored according to Paribakht and Wesche's (1997) original scoring criteria described above. Scores of 1, 2, 3, 4, or 5 were possible based on which category was marked and the extent of the answers. However, in lieu of summing up these scores and treating them as ordinal data, the scores were analyzed descriptively in order to reveal the distribution of words in each score category. This scoring method illuminated differences in depth of vocabulary knowledge at the beginning and end of treatment.

There was a total of 50 test items each on the pretest and posttest, which included 10 IT words, 10 IC words, 10 IT-IC words, and 20 distractors; only the 30 target word items were scored. With regard to the first scoring method, one rater scored the pretest and posttest twice at two time intervals. Intra-rater reliability was 100%. Regarding the second method, two raters independently scored the pretest and posttest. Interrater reliability was .99, and disagreements were resolved through discussion.

5.4. Procedure

During the treatment phase, the participants were first given 25 minutes to study a vocabulary sheet containing a list of 10 IT target words, after which the vocabulary sheet was collected. Then, they read Chapter 1 of *The Elephant Man* and then answered comprehension questions about the text. After a short break, they were given 25 minutes to study a second vocabulary sheet which contained a list of 10 IT-IC target words. Then, the vocabulary sheet was collected and participants read Chapters 2 and 3 of *The Elephant Man* and then answered comprehension questions.

Two days later, the participants read Chapters 4 and 5 of *The Elephant Man* and then answered comprehension questions. Five days later, the participants read the remaining chapters of *The Elephant Man* – Chapters 6 and 7 – and then answered the last set of comprehension questions. Participants were exposed to the IC and IT-IC target words as they read Chapters 2-7 of the reading text in which the target words had been embedded.

The pretest was administered in one sitting five days prior to the onset of treatment, while the posttest was administered in two parts in two different days to ensure that an equal interval of time, exactly nine days, passed between the end of each treatment condition and the implementation of the respective posttests. The first part of the posttest contained 25 items: 10 IT target words and 15 distractors. The second part of the posttest contained 25 items: 10 IC target words, 10 IT-IC target words, and 5 distractors. Participants completed a consent form and a background questionnaire prior to the study, and they also completed a post-study questionnaire after all assessments had concluded.

Table 1. Summary of Procedures

Day	IT	IC	**IT-IC**
1	VKS Pretest	VKS Pretest	VKS Pretest
	(5 days)	(5 days)	
6	Study Vocabulary List	Read Ch. 2 and 3	Study Vocabulary List & Read Ch. 2 and 3
8	(9 days)	Read Ch. 4 and 5	Read Ch. 4 and 5
13		Read Ch. 6 and 7	Read Ch. 6 and 7
15	VKS Posttest	(9 days)	
22		VKS Posttest	VKS Posttest

6. Results

6.1. Research Question 1

The primary research question was concerned with the effects of IT, IC, and IT-IC conditions on L2 vocabulary gains. The VKS pretest and posttest were scored dichotomously according to the number of words that were known to the participants (i.e., participants were able to provide a correct definition or synonym of the word), and participants' pretest-to-posttest gain scores were calculated for each learning condition through subtracting the pretest scores from the posttest scores. Table 2 displays the means, standard deviations, and medians of the pretest-to-posttest gain scores. The IT condition led to 10.7% of target word gains, the IC condition led to 11.3% target word gains, and the IT-IC condition led to 25.3% gains.

Table 2 Pretest to Posttest Gains

Condition	IT Gains	IC Gains	IT-IC Gains
Mean	1.07	1.13	2.53
Standard Deviations	1.2	1.68	2.06
Median	1	1	2

6.2. Research Question 2

The second research question was concerned with the differences in vocabulary gains between the three conditions. A Friedman test was run on the data that resulted from the dichotomous scoring method. The Friedman test is the non-parametric alternative to the one-way repeated measures analysis of variance (ANOVA), and was used in the current study due to the presence of outliers and unmet normality assumptions for the one-way repeated measures ANOVA.

Results showed that vocabulary gains were statistically significantly different between learning conditions, $\chi^2(2) = 16.516$, $p < .001$. Post hoc analysis, which involved pairwise comparisons with a Bonferroni correction for multiple comparisons, revealed statistically significant differences between the IT-IC (Mdn = 2) and IT (Mdn = 1) (p = 0.006) conditions and between the IT-IC and IC (Mdn = 1) (p = 0.006) conditions. In other words, the IT-IC condition led to significantly greater vocabulary gains than either the IT condition or the IC condition. There were no significant differences between the IT and IC conditions.

6.3. Research Question 3

The third research question was concerned with qualitative changes in vocabulary knowledge in relation to the three learning conditions. To investigate this question, the average number of words in each score category of the VKS (1, 2, 3, 4, and 5) at pretest and posttest were calculated. Differences between the pretest and posttest were also calculated. The results of this analysis are shown in Table 3 and displayed visually in Figure 1.

Table 3 Pretest and Posttest Mean Number of Target Words across Five VKS Score Categories

	IT			IC			IT-IC		
	Pre	Post	Pre-Post*	Pre	Post	Pre-Post*	Pre	Post	Pre-Post*
Score 1	8.3	5.23	-3.07	5.83	2.9	-2.93	6.53	2.13	-4.4
Score 2	1.47	3.4	1.93	3.53	5.33	1.8	2.73	4.57	1.84
Score 3	0.2	0.53	0.33	0.2	0.63	0.43	0.4	1.63	1.23
Score 4	0.03	0.1	0.07	0.03	0.27	0.24	0.13	0.23	0.1
Score 5	0.07	0.73	0.66	0.4	0.9	0.5	0.23	1.43	1.2

*Calculated by subtracting pretest data from posttest data.

Figure 1 Distribution of Target Words in the Five VKS Score Categories

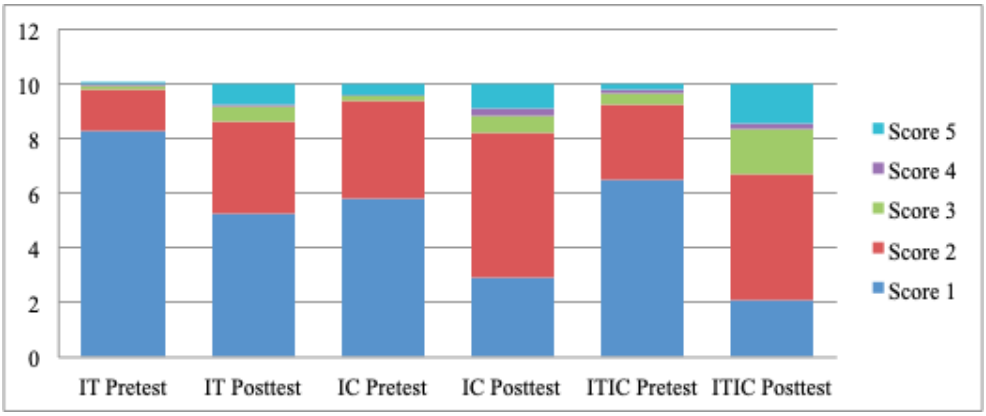


Figure 1 illustrates reductions in score category 1 and increases in score categories 2, 3, 4, and 5 from pretest to posttest for all experimental conditions. In each condition, the amount of reduction in score category 1 was greater than the amount of increase in score category 5, indicating that some words were partially acquired. In other words, some words that had started in the completely unknown category (score category 1) did not transition all the way to the full mastery category (score category 5), but they did indeed transition to a category in the middle of the scale (score categories 2, 3, or 4).

The IT-IC condition led to the greatest reduction in category 1 at 4.4 words, followed by the IT condition at 3.07 words, and then the IC condition at 2.93 words. The IT-IC, IT, and IC conditions all led to gains in about 2 words (1.84, 1.93, and 1.8, respectively) in score category 2, signifying that the form (orthography) of these words were gained, even though these words were not learned to the point of their meanings being known. As for score categories 3 and above, which signified that learners had reached the threshold of being able to associate a word's form with its meaning, the IT-IC condition led to an increase of 2.53 words (sum of increases in categories 3, 4, and 5), which was more than twice as high as the increases in the IT condition (1.06) and the IC condition (1.17). In general, this analysis corroborated the findings from the prior statistical analysis in showing that the IT-IC condition was the most effective, resulting in the greatest reduction of words (4.4) in the completely unknown category (VKS score category 1) from pretest to posttest, and the largest gains (2.53) in words whose meanings became known.

7. Discussion

The current study set out to investigate the relative efficacy of incidental, intentional, and combined intentional-plus-incidental conditions on adult ESL learners' L2 vocabulary acquisition. Results showed that the combined (IT-IC) condition resulted in the highest vocabulary gains at 25.3%, followed by the incidental-only (IC) condition at 11.3% and then the intentional-only (IT) condition at 10.7%. The IT-IC condition resulted in statistically significantly higher gains than either the IC or IT conditions; no statistically significant differences were detected between the IC and IT conditions. All treatment conditions led to partial vocabulary gains. The IT-IC condition led to the greatest reduction in words that were completely unknown, and the greatest increase in words that became known to the point that form-meaning connections could be made. Overall, the findings of the current study offer empirical support for the claim that using both intentional and incidental learning is more beneficial for L2 vocabulary acquisition than using only one of the approaches (Hunt & Beglar, 1998, 2005; Nation, 2001, Schmitt, 2010).

The finding that the IT-IC condition was more beneficial for L2 vocabulary learning than the IT condition indicates that incidental learning that follows intentional learning can lead to greater retention. While both the IT and IT-IC conditions had a deliberate word study component, attrition in word knowledge was most likely lessened in the IT-IC condition due to additional encounters with words in context during reading, which helped to consolidate and enhance the knowledge of newly learned words. This finding aligns with the usage-based perspective which maintains that implicit learning occurs as learners' perceptual system take tallies of form-meaning associations when exposed to language input (Ellis, 2013). With each additional encounter, associations between form and meaning are strengthened (Goldberg, 2006). In turn, the finding that the IT-IC condition was also more advantageous than the IC condition for L2 vocabulary development indicates that intentional learning used prior to incidental learning can enhance the efficacy of subsequent learning. For example, although both the IC and IT-IC conditions exposed participants to target words only 3 or 8 times during the reading sessions, even words met as few as 3 times under the IT-IC condition showed improvement in knowledge, which reveals another potential advantage of IT – that it may speed up learning or compensate for lack of exposure. This is notable, given that previous research has shown that at least 8 to 10 encounters are needed for learners to have a reasonable chance of learning new words during reading (Horst, 2005; Pellicer-Sánchez, 2015; Pellicer-Sánchez & Schmitt, 2010; Webb, 2007). The effectiveness of encountering words during reading in the IT-IC condition was most likely enhanced because it occurred after a session of intentional learning during which form-meaning associations were initiated. This finding also aligns with

the usage-based perspective which maintains that explicit knowledge of novel form-meaning constructions can aid in subsequent implicit learning of these constructions (Ellis, 2006, 2013).

The finding that there were no statistically significant differences between the IC and IT conditions contradicts the findings of the majority of previous research that has shown that intentional learning is generally superior to incidental learning (Schmitt, 2008; Won, 2008). The reason for the low vocabulary gains seen in previous ICVL studies are often attributed to the fact that L2 learners often fail to notice many unknown words in the texts they read (Laufer & Yano, 2001) and encounter difficulty correctly inferring the meanings of unknown words when contextual clues for word meanings are non-existent, unhelpful, or misleading (Beck, McKeown, & McCaslin, 1983; Holmes & Ramos, 1993; Schatz & Baldwin, 1986). On the other hand, the reason for the high vocabulary gains seen in previous ITVL studies are often attributed to the fact that the noticing of unknown words and the accessing of their correct meanings are much more likely to occur. In addition, Eagle and Leiter (1964) have pointed out that learners who have an intention to learn will most likely engage in constructing some kind of plan to guide memorization and recall, and that this act alone will entail greater cognitive processing compared to when learners start out without any such plan (see also discussion on the Involvement Load Hypothesis in Hulstijn & Laufer, 2001). Despite these factors, the reason for the different finding in the current study may be due to the delayed implementation (nine days after the end of treatment) of the vocabulary posttests. As noted earlier, a weakness of using intentional learning alone is the tendency for steep attrition of vocabulary knowledge (e.g., Laufer, 2005; Nakata, 2008). The reason why intentionally acquired words are so amenable to attrition may be because the form-meaning links of newly acquired words exist as isolated bits of knowledge and are stored in short-term memory rather than being integrated into a larger system of preexisting networks stored in long-term memory. Without reinforcement, this type of knowledge is fragile and will be easily forgotten with the passage of time.

In fact, scores on vocabulary posttests that are administered several days after treatment may actually be a more realistic measure of vocabulary gains than posttests that are administered immediately after treatment. These immediate posttests, which fail to account for attrition, may represent vocabulary knowledge stored temporarily in learners' short-term memory rather than vocabulary knowledge that has been integrated into learners' long-term memory. A common practice in L2 research is to employ both immediate and delayed posttests, but a caveat to this method is that the immediate posttest provides a learning experience which can affect the performance on the succeeding posttests. Psychological research (e.g., Johnson & Mayer, 2009;

McDaniel, Anderson, Derbish, & Morrisette, 2007; Roediger & Karpicke, 2006) has shown that the so-called "testing effect" is formidable in leading to increases in learning, beyond the benefits of simply providing additional study exposure.

Apart from the timing of the posttest, another reason for the current study's finding that the IC and IT conditions had no significant differences may be due to the difference in time duration and distribution for each learning condition. While the IC condition was distributed over three days in three sessions, the IT condition took place in one day in one 25-minute session. These different time durations and distributions reflect the typical characteristics of the respective pedagogical tasks; on the whole, intentional learning processes (e.g., which occur while studying words on a list) tend to be massed and intensive while incidental learning processes (e.g., which occur while reading for pleasure) tend to be distributed and relatively less intensive. These differences in time frames may have played a tangible role in determining the relative effectiveness of incidental and intentional learning observed in the present study. Psychological research has shown that while massing practice tends to produce better performance on immediate posttests, distributing practice over a longer period of time leads to better longer term performance and recall (Bjork, 1994; Murray & Undermann, 2003; Sobel, Cepeda, & Kapler, 2011).

Another notable finding from the present study is that gains were quite low across the treatment conditions. Even the IT-IC condition only led to 25.3% gains. The reason may be that the treatment tasks failed to elicit deep processing of target words. As stated by Schmitt (2008), "the overriding principle for maximizing vocabulary learning is to increase the amount of engagement learners have with lexical items" (p. 329). This notion originates from Craik and Lockhart's (1972) psychological theory of human memory on Levels of Processing (LOP), which posits that the amount of learning is dependent on the differing levels, ranging from shallow to deep, at which information is processed. According to this theory, rich, deep processing of information leads to greater learning than shallow processing. Hence, the efficacy of incidental, intentional, and combined intentional and incidental learning conditions on L2 vocabulary acquisition is likely contingent on the extent to which these conditions lead to learners' engagement (i.e., involvement, depth of processing) with vocabulary words, a topic for future research.

Interestingly, despite the generally low gains, all learning conditions led to modest increases in depth of vocabulary knowledge, as shown in the VKS score distribution analysis. For example, all treatment conditions led to gains in about 2 words (1.84, 1.93, and 1.8, respectively) in score category 2. The transition from VKS score category 1 to category 2 represents the beginning stages in vocabulary learning, which occurs

when a learner encounters and *notices* a word's form (written or spoken) and registers it in memory (Schmidt, 1990). It seemed that the acquisition of word form (orthography) occurred relatively quickly and with just a small amount of exposure compared to other aspects of word knowledge (e.g., form-meaning link). This finding is consistent with previous studies of incidental vocabulary acquisition (e.g., Pellicer-Sánchez & Schmitt, 2010; van Zeeland & Schmitt, 2013) which have shown that vocabulary learning interventions typically result in the greatest gains in word form, followed by gains in grammatical aspects (i.e., part of speech), followed lastly by gains in form-meaning connections. Although often overlooked, this preliminary step of noticing and registering the word form is significant, given that words whose written forms are noticed and registered in an initial encounter are likely to be learned faster and to greater depths when they are encountered subsequently in a new environment.

Our results also show that the IT-IC condition was the only condition that led to the acquisition of two to three words (2.53) that was learned well enough for learners to be able to link their forms and meanings; the IT and IC conditions both led to only about one word that was learned to this level of depth (1.06 and 1.17, respectively). Moreover, the IT-IC condition led to about one word (1.2) that could be used successfully in a sentence context, with both semantic and grammatical accuracy, while the IT and IC conditions both showed far less gains in this regard (0.66 and 0.5, respectively). The fact that not all words whose meanings were identified correctly were also used correctly within a sentence illuminates that for beginning and intermediate L2 learners, knowing the form-meaning link of a word does not automatically lead to the ability to use the word, not even in a sentential context. Knowledge of semantic and grammatical functions of a word may reflect a depth of vocabulary knowledge that is indeed more difficult to acquire. L2 learners may thus benefit from explicit instruction on these subtle word knowledge aspects in addition to continued exposure to these aspects in incidental learning contexts.

It must be recognized that the current study has several limitations. For one, due to the small sample size, a within-subjects research design was used, which entails a threat to validity with respect to "multiple-treatment interference," the possibility of the participation in one condition affecting performance in the other conditions (Wiersma & Jurs, 2009: 141). Other limitations to the current study are related to the small number of target words and the lack of control of relevant variables such as word features, context in which words appeared. Although the target words were randomly assigned to treatment conditions to nullify systematic differences, it is possible that their intrinsic lexical properties – semantic complexity, length of word form, and the contexts in which the IC and IT-IC words appeared in the reading text – confounded the effects of treatment. Finally, because time-on-task for the IC, IT and

IT-IC conditions was not controlled in the current study, it is unknown how these interventions would have affected L2 vocabulary development, had they taken place within the same time frame.

In closing, although it is widely assumed that integrating intentional and incidental learning conditions is beneficial for fostering L2 lexical development, research has yet to "discover the ways in which such a combination [of incidental and intentional approaches] can be put into practice most effectively considering class-time limitations and classroom resources that L2 teachers and learners have at their disposal" (Chacón-Beltrán, Abello-Contesse, Torreblanca-López, 2010: 6). Although the current study was an initial step toward this endeavor, future research that explores different *combinations* of incidental and intentional learning tasks - for example, using them in alternation, reversed order, or within different time frames - and different *types* of incidental and intentional learning tasks – for example, listening tasks for incidental learning, focused word exercises for intentional learning – will lead to a better understanding of how incidental and intentional conditions can be optimally integrated to aid L2 vocabulary acquisition. Moreover, longitudinal studies that examine the effects of long-term vocabulary interventions will shed light on how the incremental process of vocabulary acquisition is affected by incidental, intentional, and combined approaches over time.

8. References

Anderson, J. R. (1982). *Acquisition of cognitive skill. Psychological Review, 89*(4), 369-406.

Adolphs, S., & Schmitt, N. (2003). Lexical coverage of spoken discourse. *Applied Linguistics, 24*(4), 425-438.

Ahmad, J. (2011). Intentional vs. incidental vocabulary learning. *Interdisciplinary Journal of Contemporary Research in Business, 3*(5), 67-75.

Barcroft, J. (2004). Second language vocabulary acquisition: A lexical input processing approach. *Foreign Language Annals, 37*(2), 200-208.

Barcroft, J. (2009). Effects of synonym generation on incidental and intentional L2 vocabulary learning during reading. *TESOL Quarterly, 43*(1), 79-103.

Barcroft, J., & Rott, S. (2010). Partial word form learning in the written mode in L2 German and Spanish. *Applied Language Learning, 31*(5), 623-650.

Beck, I. L., McKeown, M. G., & McCaslin, E. S. (1983). Vocabulary development: All contexts are not created equal. *The Elementary School Journal, 83*(3), 177-181.

Bjork, R. A. (1994). Memory and metamemory considerations in the training of human beings. In J. Metcalf & A. P. Shimamura (Eds.), *Metacognition: Knowing about knowing* (pp. 185-205). Cambridge, MA: MIT Press.

Bruton, A., Garcia Lopez, M., & Esquiliche Mesa, R. (2011). Incidental L2 vocabulary learning: An impracticable term? *TESOL Quarterly, 45*(4), 759-768.

Chacón-Beltrán, R., Abello-Contesse, C., Torreblanca-López, M.M. (2010). Vocabulary teaching and learning: Introduction and overview. In R. Chacón-Beltrán, C. Abello-Contesse, M. Torreblanca-López (Eds.), *Insights into Non-native Vocabulary Teaching and Learning* (pp. 1–12). Bristol, Blue Ridge Summit: Multilingual Matters.

Craik, F. I. M., & Lockhart, R. S. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior, 11*(6), 671-684.

Daskalovska, N. (2011). The impact of reading on three aspects of word knowledge: spelling, meaning, and collocation. *Procedia Social and Behavioral Sciences, 15*(2011), 2334-2341.

Day, R. R., Omura, C., & Hiramatsu, M. (1991). Incidental EFL vocabulary learning and reading. *Reading in a Foreign Language, 8*, 689-696.

DeKeyser, R. (2015). Skill acquisition theory. In B. VanPatten & J. Williams (Eds.), *Theories in second language acquisition: An introduction* [2nd Edition] (pp. 94 – 112). New York, NY: Routledge.

Dupuy, B., & Krashen, S. D. (1993). Incidental vocabulary acquisition in French as a foreign language. *Applied Language Learning, 4*(1-2), 55-63.

Eagle, M., & Leiter, E. (1964). Recall and recognition in intentional and incidental learning.

Ellis, R. (1994). Factors in the incidental acquisition of second language vocabulary from oral input: A review essay. *Applied Language Learning, 5*(1), 1-32.

Ellis, N. C. (2006). Cognitive perspectives on SLA: The Associative-Cognitive CREED. *AILA Review, 19*, 100-121.

Ellis, N. (2013). Second language acquisition. In G. Trousdale & T. Hoffman (Eds.), *Oxford handbook of construction grammar* (pp. 365-378). Oxford: Oxford University Press.

Ellis, N., & Wulff, S. (2015). Usage-based approaches to SLA. In B. VanPatten & J. Williams (Eds.), *Theories in second language acquisition: An introduction* [2nd Edition] (pp. 75 – 93). New York, NY: Routledge.

Ertürk, Z. Y. (2016). The effect of glossing on EFL learners' incidental vocabulary learning in reading. *Procedia - Social and Behavioral Sciences, 232*(2016), 373-381.

File, K. A., & Adams, R. (2010). Should Vocabulary Instruction Be Integrated or Isolated? *TESOL Quarterly, 44*(2), 222-249.

Folse, K. S. (2006). 6). The effect of type of written exercise on L2 vocabulary retention. *TESOL Quarterly, 40*, 273-293.

Gass, S. (1999). Discussion: Incidental vocabulary learning. *Studies in Second Language Acquisition, 21*(2), 319-333.

Goldberg, A. E. G. (2006). Learning linguistic patterns. *Psychology of Learning and Motivation, 47*, 33-63.

Hill, M., & Laufer, B. (2003). Type of task, time-on-task and electronic dictionaries in incidental vocabulary acquisition. *International Review of Applied Linguistics in Language Teaching, 41*(2), 87-106.

Holmes, J., & Ramos, R. G. (1993). False friends and reckless guessers: Observing cognate recognition strategies. In T. Huckin, M. Haynes & J. Coady (Eds.), *Second language reading and vocabulary learning* (pp. 86-107). Norwood, NJ: Ablex.

Horst, M. (2005). Learning L2 vocabulary through extensive reading: A measurement study. *Canadian Modern Language Review, 61*, 355-382.

Horst, M., Cobb, T., & Meara, P. (1998). Beyond *A Clockwork Orange*: Acquiring second language vocabulary through reading. *Reading in a Foreign Language, 11*, 207-223.

Hulstijn, J. (1992). Retention of inferred and given word meanings: experiments in incidental vocabulary learning. In P. J. Arnaud & H. Bejoint (Eds.), *Vocabulary and Applied Linguistics* (pp. 113-125). London: Macmillan.

Hulstijn, J. H. (2001). Intentional and incidental second language vocabulary learning: A reappraisal of elaboration, rehearsal and automaticity. In P. Robinson (Ed.), *Cognition and second language instruction* (pp. 258-286). Cambridge: Cambridge University Press.

Hulstijn, J. H. (2003). Incidental and intentional learning. In C. J. Doughty & M. H. Long (Eds.), *The handbook of second language acquisition* (pp. 349-381). Oxford: Blackwell.

Hulstijn, J. H., Hollander, M., & Greidanus, T. (1996). Incidental vocabulary learning by advanced foreign language students: The influence of marginal glosses, dictionary use, and reoccurrence of unknown words. *The Modern Language Journal, 80*(3), 327-339.

Hulstijn, J., & Laufer, B. (2001). Some empirical evidence for the involvement load hypothesis in vocabulary acquisition. *Language Learning, 51*, 539-558.

Hunt, A., & Beglar, D. (1998). Current research and practice in teaching vocabulary. *The Language Teacher, 22*(1). http://www.jalt.publications.org/tlt/articles/1998/01/hunt

Hunt, A., & Beglar, D. (2005). A framework for developing EFL reading vocabulary. *Reading in a Foreign Language, 17*(1), 23-59.

Johnson, C., & Mayer, R. (2009). A testing effect with multimedia learning. *Journal of Educational Psychology, 101*(3), 621-629.

Knight, S. (1994). Dictionary: The tool of last resort in foreign language reading? A new perspective. *Modern Language Journal, 78*, 285-299.

Krashen, S. (1987). *Principles and practice in second language acquisition*. New York, NY: Prentice Hall.

Krashen, S. (1989). We acquire vocabulary and spelling by reading: Additional evidence for the input hypothesis. *Modern Language Journal, 73*(4), 440-464.

Krashen, S. (2004). *The power of reading*. Portsmouth, NH: Heinemann.

Krashen, S. (2013). Reading and vocabulary acquisition: Supporting evidence and some objections. *Iranian Journal of Language Teaching Research, 1*(1), 27-43.

Kweon, S.-O., & Kim, H.-R. (2008). Beyond raw frequency: Incidental vocabulary acquisition in extensive reading. *Reading in a Foreign Language, 20*(2), 191-215.

Laufer, B. (2005). Focus on form in second language vocabulary acquisition. In S. H. Foster-Cohen, M. P. Garcia-Mayo & J. Cenoz (Eds.), *EUROSLA Yearbook 5* (pp. 223–250). Amsterdam: Benjamins.

Laufer, B., & Yano, Y. (2001). Understanding unfamiliar words in a text: do L2 learners understand how much they don't understand. *Reading in a Foreign Language, 13*(2), 549-566.

Luppescu, S., & Day, R. R. (1993). Reading, dictionaries, and vocabulary learning. *Language Learning, 43*(2), 263-279.

Meara, P. (1990). A note on passive vocabulary. *Second Language Research, 6*(2), 150-154.

McDaniel, M, Anderson, J., Derbish, M., & Morrisette, N. (2007). Testing the testing effect in the classroom. *European Journal of Cognitive Psychology, 19*(4-5), 494-513.

Murray, S. R., & Undermann, B. E. (2003). Massed versus distributed practice: Which is better? *CAHPERD Journal, 28*(1), 19-22.

Nakata, T. (2008). English vocabulary learning with word lists, word cards and computers: Implications from cognitive psychology research for optimal spaced learning. *ReCALL, 20*(1), 3–20.

Nakata, T. (2017). Does repeated practice make perfect? The effects of within-session repeated retrieval on second language vocabulary learning. *Studies in Second Language Acquisition*, *39*(4), 653–679.

Nation, I. S. P. (2001). *Learning vocabulary in another language*. Cambridge: Cambridge University Press.

Nation, I. S. P. (2011). Research into practice: Vocabulary. *Language Teaching*, *44*(4), 529–539.

Paradis, M. (1994). Neurolinguistic aspects of implicit and explicit memory: Implications for bilingualism and second language acquisition. In N. Ellis (Ed.), *Implicit and Explicit Language Learning* (pp. 393-419). London: Academic Press.

Paribakht, T. S. (2005). The influence of L1 lexicalization on L2 lexical inferencing: A study of Farsi-speaking EFL learners. *Language Learning, 55*(4), 701-748.

Paribakht, T. S., & Wesche, M. (1997). Vocabulary enhancement activities and reading for meaning in a second language. In J. Coady & T. Huckin (Eds.), *Second language vocabulary acquisition* (pp. 174-200). Cambridge: Cambridge University Press.

Pellicer-Sánchez, A. (2015). Incidental L2 vocabulary acquisition from and while reading: An eye-tracking study. *Studies in Second Language Acquisition, 38*(1), 1-34.

Pellicer-Sánchez, A., & Schmitt, N. (2010). Incidental vocabulary acquisition from an authentic novel: Do Things Fall Apart? *Reading in a Foreign Language, 22*(1), 31-55.

Pigada, M., & Schmitt, N. (2006). Vocabulary acquisition from extensive reading: A case study. *Reading in a Foreign Language, 18*, 1-28.

Pitts, M., White, H., & Krashen, S. (1989). Acquiring second language vocabulary through reading: A replication of the Clockwork Orange study using second language acquirers. *Reading in a Foreign Language, 5*, 271-275.

Postman, L., & Adams, P. A. (1956). Studies in incidental learning: IV. The interaction of orienting tasks and stimulus materials. *Journal of Experimental Psychology, 51*(5), 329.

Prince, P. (2012). Towards an instructional programme for L2 vocabulary: Can a story help? *Language Learning & Technology, 16*(3), 103-120.

Qian, D. D. (1996). ESL vocabulary acquisition: contextualization and decontextualization. *The Canadian Modern Language Review, 53*(1), 120-142.

Read, J. (2000). *Assessing vocabulary*. Cambridge: Cambridge University Press.

Read, J. (2004). Research in teaching vocabulary. *Annual Review of Applied Linguistics, 24*, 146-161.

Rieder, A. (2003). Implicit and explicit learning in incidental vocabulary acquisition. *VIEWS, 12*, 24-39.

Roediger, H. L., & Karpicke, J. D. (2006). Test-enhanced learning: Taking memory tests improves long-term retention. *Psychological Science*, *17*(3), 249-255.

Schatz, E. K., & Baldwin, R. S. (1986). Context clues are unreliable predictors of word meanings. *Reading Research Quarterly, 21*(4), 439-453.

Schmidt, R. (1990). The role of consciousness in second language learning. *Applied Linguistics, 11*, 129-158.

Schmidt, R. (1994). Deconstructing consciousness in search of useful definitions for applied linguistics. In J. Hulstijn and R. Schmidt (Eds.), *Consciousness in Second Language Learning. AILA Review, 11*, 11-26.

Schmitt, N. (2008). Review article: Instructed second language vocabulary learning. *Language Teaching Research, 12*(3), 329-363.

Schmitt, N. (2010). *Researching vocabulary: A vocabulary research manual*. New York: Palgrave Macmillan.

Sobel, H. S., Cepeda, N. J., & Kapler, I. V. (2011). Spacing effects in real-world classroom vocabulary learning. *Applied Cognitive Psychology, 25*(5), 763-767.

Sonbul, S., & Schmitt, N. (2010). Direct teaching of vocabulary after reading: Is it worth the effort? *ELT Journal, 64*(3), 253-260.

Thorndike, E. (1908). Memory for paired associates *Psychological Review, 15*, 122-138.

Van Zeeland, H., & Schmitt, N. (2013). Lexical coverage in L1 and L2 listening comprehension: The same or different from reading comprehension? *Applied Linguistics, 34*(4), 457-479.

Vicary, T. (1989). *The elephant man*. Oxford, England: Oxford University Press.

Waring, R., & Nation, I. S. P. (2004). Second language reading and incidental vocabulary learning. *Angles on the English-Speaking World, 4*, 11-23.

Waring, R., & Takaki, M. (2003). At what rate do learners learn and retain new vocabulary from reading a graded reader? *Reading in a Foreign Language, 15*(2), 130-163.

Webb, S. (2007). The effects of repetition on vocabulary knowledge. *Applied Linguistics, 28*(1), 46-65.

Webb, W. B. (1962). The effects of prolonged learning on learning. *Journal of Verbal Learning and Verbal Behavior, 1*, 173-182.

Wesche, M., & Paribakht, T. S. (1999). Introduction. *Studies in Second Language Acquisition, 21*(2), 175-180.

Wesche, M., & Paribakht, T. S. (2009). *Lexical inferencing in first and second language: Cross-linguistic dimensions*. Clevedon: Multilingual Matters.

Wiersma, W., & Jurs, S. G. (2009). *Research methods in education: An introduction.* Pennsylvania: Pearson/Allyn and Bacon.

Won, M. (2008). *The effects of vocabulary instruction on English language learners: A meta-analysis.* (Ph.D. dissertation), Texas Tech University.

Zimmerman, C. (1997). Do reading and interactive vocabulary instruction make a difference? An empirical study. *TESOL Quarterly, 31*, 121-140.

Developing interactional repertoires in the classroom through dynamic strategic interactio scenarios

Rémi A. van Compernolle
vancomp@cmu.edu

Nuria Ballesteros Soria
nballest@andrew.cmu.edu
Department of Modern Languages
Carnegie Mellon University

Abstract

In this article, we report on the implementation of Dynamic Strategic Interaction Scenario tasks as one approach to developing L2 learners' interactional repertoires. The tasks involve pre-task planning, a performance, and immediate feedback during a whole-class debriefing discussion. We focus specifically on the appropriation of turn allocation devices within a single class meeting in which three small groups performed the same scenario. We show how the first group's performance prompted a focus on turn allocation during their debriefing, and how the subsequent groups were able to build on the feedback in their own performances. We discuss our findings and their implications for research and pedagogy along three dimensions: 1) the role of feedback as mediation in the debriefing discussions; 2) the contribution of task repetition from a group-as-collective perspective; and 3) the documentation of interactional repertoire development over time.

Keywords: Interactional repertoires; Interactional competence; Dynamic strategic interaction scenarios; Turn allocation; Spanish as a second or foreign language

Resumen

Este este artículo se presenta la implementación de escenarios de interacción estratégica y dinámica como un método para promover el desarrollo de los recursos de interacción entre discentes de lenguas extranjeras y de segundas lenguas. Las tareas constan de una fase de planificación y una de representación, seguidas de una discusión y puesta en común en asamblea en la que se ofrece retroalimentación inmediata. El presente artículo se centra especialmente en la apropiación de recursos para la cesión

del turno de habla durante una clase en la que tres pequeños grupos representan el mismo escenario. Se muestra cómo la representación del escenario por parte del primer grupo motiva una conversación sobre la cesión del turno de habla durante la puesta en común y cómo el resto de grupos incorpora la retroalimentación en sus respectivos escenarios. Los resultados y las implicaciones pedagógicas y de investigación se abordan en relación con tres temas: 1) el papel mediador de la retroalimentación durante las discusiones de las puestas en común; 2) las aportaciones de la repetición de tareas desde una perspectiva grupal y colectiva y 3) el estudio del desarrollo del repertorio de recursos de interacción a lo largo de la sesión.

Palabras clave: repertorios de recursos de interacción; competencia interaccional; escenarios de interacción estratégica y dinámica; cesión del turno de habla; español como segunda lengua o como lengua extranjera

1. Introduction

Developing speaking abilities is an important goal for second and foreign language (L2 for simplicity) education. Informed by models of communicative competence (Bachmann, 1990; Canale & Swain, 1980; Celce-Murcia, 2007; Celce-Murcia, Dörynei, & Thurrell, 1995), pedagogical arrangements aimed at developing learners' speaking abilities have traditionally focused on the learner's capacity to generate and understand spoken language as an individual competence. However, more recent approaches to understanding L2 speaking in terms of interactional competence (IC) (Hall, Hellermann, & Pekarek Doehler, 2011; Salaberry & Kunitz, 2019; Waring, 2018) and the deployment of interactional repertoires (Hall, 2018) have refocused attention on the way in which speaking is coproduced between interlocutors. This perspective has broadened the purview of speaking development to include not only linguistic features of communication (e.g., grammar, lexis, speech acts) but more importantly the interactive practices that underly turn-taking systems, conversational repair, and action sequencing.

In this article, we report findings from an initial attempt at supporting the development of L2 interactional repertoires through a series of dynamic strategic interaction scenario (DSIS) tasks (van Compernolle, 2014a, 2014b, 2018) that were integrated into an existing elementary-level Spanish course at a US university. The DSIS tasks were performed by multiple small groups in front of the class, and feedback on language use and interactional strategies was provided by the teacher and other students after each performance. Our focus in the present article is on the microgenetic development of turn allocation practices that occurred in one class meeting, where each group was able to build on the performances of and feedback given to prior groups.

2. Background

2.1. Interactional competence and interactional repertoires in L2 learning

The notion of interactional competence as a pedagogical target in L2 settings dates back at least to Kramsch's (1986) critical appraisal of the proficiency movement in language teaching and testing circles. Kramsch points out that proficiency models, grounded as they are in individualistic conceptions of communicative competence (Canale & Swain, 1980), assume that "language teaching and learning are input-output processes, guided by a linear acquisition of grammatical structures and that the major criterion of accuracy is grammar correction" (p. 370). This assumption, she argues, ignores the fact that successful communication involves an interaction between interlocutors as they negotiate the meaning and significance of language-in-use. While more recent models of communicative competence (Celce-Murcia, 2007) and speaking proficiency (ACTFL, 2012) have recognized the role of such interactive phenomena as turn-taking and meaning negotiation, they remain centered around the individual rather than the processes and outcomes of communication between people.

In the 1990s and early 2000s, research informed by linguistic anthropology and conversation analysis (e.g., Hall, 1993, 1995; He & Young, 1998; Young, 2008) began to take seriously the idea that an L2 user's competence during communication was an interactive phenomenon. In this line of inquiry, interactional competence is not something acquired and possessed by the individual; instead, it is what becomes relevant and available in one's interactions with others. This is to say that competence is distributed across interlocutors. It is contingent not only on lexicogrammatical forms but on shared understandings of the context as well as shared orientations to relevant next actions in interactive discourse.

As research in this domain has gained attention among L2 researchers, however, there has been a conflation of two leading concepts, as Hall (2018) points out. On one hand, interactional competence has been used to refer to the "basic interaction infrastructure of human sociality" (p. 28), a notion derived from conversation analysis. On the other hand, the term has been used to describe "variability of individual knowledge within and across social groups" (p. 28), which is the focus of work in linguistic anthropology. While both traditions have come to inform interactional competence scholarship, Hall argues that we need an alternative concept to describe the types of interactive practices that L2 learners appropriate in their development—*interactional repertoires*. This terminology underscores the idea that interactional *competence* is a kind of universal underlying infrastructure for human sociality whereas in L2 development the task for learners is to appropriate a diverse and context-sensitive

repertoire of interactive resources for engaging in communicative activity. The resources that make up a learner's interactional repertoire constitute the objects of analysis in this line of inquiry.

Research on instructional arrangements specifically designed to develop and expand learners' interactional repertoires is just now emerging (e.g., Barraja-Rohan, 2011; Kunitz & Yeh, 2019; Lilja & Piirainen-Marsh, 2019; Waring, 2019). Although the contexts of instruction, languages, and foci vary across these studies, three common threads run through them. First, the focus of pedagogy must shift from lexicogrammatical forms to interactive practices, especially with regard to turn taking strategies. Second, awareness of interactive practices should be fostered so that learners can consciously attend to and reflect on the appropriateness of their interactive practices. Third, learners need opportunities to participate in interactions where they can enact their developing interactive repertoires. Thus far, research has predominately focused on organizing pedagogical arrangements around awareness-raising tasks, following Barraja-Rohan's (2011) suggestion that language learners become, to some extent at least, conversation analysts in their own right. This means introducing some of the central concepts of turn-taking from the conversation analysis (CA) literature (Sacks, Schegloff, & Jefferson, 1974) and having learners analyze speech samples, such as recordings and transcriptions of native speaker speech events. To our knowledge, no research has examined the integration of speaking tasks specifically designed to develop both awareness of interactional practices and learner IC in a classroom context.

2.2. Dynamic strategic interaction scenarios

Dynamic strategic interaction scenarios (DSISs) were originally developed as part of a Vygotskian approach to teaching pragmatics through concept-based instruction as a means of focusing learners on both meaning and form in communicative tasks (van Compernolle, 2014a, 2014b, 2018). The tasks draw on DiPietro's (1987) strategic interaction approach to L2 instruction that involves three stages: (i) a rehearsal, where learners can reflect on and plan appropriate language for (ii) a performance, during which the scenario is executed, which is followed by (iii) a debriefing in which the teacher and other students can provide feedback. In van Compernolle (2014a, 2014b), DSISs were carried out in one-on-one tutoring sessions during which the tutor was able to provide assistance, or mediation, during the performance following insights from dynamic assessment (Poehner, 2008). In this way, the tasks played a dual role as assessments of learner abilities and instructional interventions. As discussed in van Compernolle (2018), however, an extension to the classroom would involve whole group feedback as mediation during the debriefing stage. The idea would be to engage

in a group dynamic assessment (Poehner, 2009) in which small groups of students could build on the performances of and feedback provided to previous groups.

2.3. Turn allocation

A central focus of pedagogical arrangement designed to enhancing L2 interactional repertoires is the turn-taking system (Barraja-Rohan, 2011). Drawing on conversation analytic research starting in the 1970s (Sacks et al., 1974; see Schegloff, 2007; Sidnell, 2011 for more recent work), the claim is that L2 IC is grounded in the ability to negotiate turns-at-talk. One important dimension of turn-taking is how turns are allocated (i.e., selecting the next speaker). As outlined in Sacks et al. (1974), there are three hierarchically organized options for allocating next turns. First, the current speaker may select the next speaker explicitly (e.g., by name) or implicitly (e.g., by gaze, gesture, or context and content of speech), who in turn has both the right and obligation to take the floor. Second, if no next speaker is selected by the current speaker, other participants can self-select (e.g., to respond to an open question or to proffer a new topic if the preceding turn closes a prior sequence). Third, the current speaker may elect to continue his or her turn if no other participant self-selects as next speaker.

The importance of turn-allocation mechanisms becomes especially clear in interactions involving three or more participants because no single currently non-speaking party can presume to be the next speaker. In other words, since the next speaker could be any one of two or more non-speaking participants, there is the potential for competition in being allocated or self-selecting to take up the next turn. As Sacks et al. (1974) write:

a current non-speaker, if interested in speaking next, will be under constraint to self-select at first possible transition point, and at each successive such point. Furthermore, if a current speaker is interested in choosing among potential next speakers, he will be under constraint to accomplish the selection before first possible transition place . . . , lest an undesired current non-speaker self-select at that point. (Sacks et al., 1974, p. 713)

Thus, learning to allocate turns to a desired next speaker and to self-select as next speaker is key to developing the capacity to manage turn-taking in interaction.

Following Hall's (2018) proposal to shift toward thinking in terms of interaction repertoires, we conceptualize turn allocation itself as part of the underlying infrastructure of competent interaction. By contrast, the appropriation of specific turn allocation devices—the actual practices by which turn allocation is achieved—

constitutes the development of one's interactional repertoire. Our interest as analysts, then, is documenting the expansion of learners' repertoires for allocating next turns.

3. Methods

3.1. Context and participants

The data we draw from for this article were collected as part of a larger project examining speaking development in an elementary-level US university Spanish classroom. The class was audio/videorecorded throughout the academic term, typically two times per week starting in week 4 of the term, resulting in approximately 18 hours of data. The class was taught by the second author of this paper, a native speaker of Spanish. The first author supervised the study and collected the data. There were 13 university students enrolled in the class, all of whom consented to participate in the study.

3.2. Implementation of DSIS tasks

DSIS tasks were integrated into instruction as a way of complementing the existing curriculum, with an eye toward enhancing the development of learners' interactional repertoires. In other words, DSISs were designed to follow and build on the themes, grammar, and vocabulary presented in the textbook. The DSISs were assigned to small groups, typically 2-3 students, each of whom had a different role to perform in the scenario and conflicting agendas that they would need to negotiate. Following DiPietro (1987), while all students understood the general context of the scenario, they were not made aware of the specific details of each other's roles.

DSISs were completed in three stages following van Compernolle's (2018) recommendations for the classroom. In the first stage, students were provided with a description of the scenario and information about the role they would play. As homework, they were instructed to reflect on the language and interactional strategies that would be useful in the scenario (see Appendices A, B, and C). In the subsequent class meeting, the students met briefly with students from other groups who had been assigned the same role so that they could share and compare their homework, strategize about useful language and interactional resources, and so on. The second stage was the performance. Each group performed the scenario in front of the class, which lasted approximately 3-4 minutes each. The third stage was the debriefing or feedback stage. After each group's performance, the instructor and students provided an evaluation of the performance and offered suggestions for improvement. In other

words, feedback (stage 3) way layered between each group's performance (stage 2) so that each subsequent group had the chance to benefit from the feedback provided to prior groups as a form of cumulative group dynamic assessment (Poehner, 2009).

3.3. Analytic procedures

For this article, we have decided to illustrate the potential benefits of cumulative feedback on DSIS performances during a single class period. To this end, we began our analysis by reviewing the audio-/video-recordings made during DSIS performances (n = 3) and identifying possible foci of interest (e.g., emergence of learning opportunities, feedback strategies). The current analysis centers on the emergence of learning opportunities around turn allocation that emerged during the second DSIS task. We transcribed the performances and feedback discussions. Transcription conventions are provided in Appendix D. We then identified the interactive practices used for allocating next turns in each of the performances as well as the practices, suggestions, and so on that were discussed in the feedback stages. In doing so, we were able to reconstruct the microgenetic development of turn allocation practices across multiple groups during a single class meeting. In the interest of space, we have limited our analysis to the first three group performances and feedback stages.

4. Findings and analysis

We report our findings in three parts. First, we present Group 1's scenario performance and feedback stages, focusing on turn allocation practices, which emerged as a topic in the feedback discussion. Second, we show Group 2's scenario performance, in which the participants incorporated some of the feedback from the prior group's performance, as well as their feedback stage in which the students and teacher expanded their discussion of turn allocation practices. Third, we analyze Group 3's performance to show how they built on the prior two groups' performances and the class's feedback.

4.1. The emergence of turn allocation as an instructional focus

Excerpt 1 shows the first minute or so of Group 1's performance (Greg, Frank, and Tianyu; all names are pseudonyms). Greg opens the interaction by initiating a greeting sequence that Frank and Tianyu both participate in (lines 1-7). At line 8, then, Greg responds to Frank's prior turn, *más o menos* (line 7), with a follow up question, *¿Por qué?*, thus allocating the next turn to Frank, who goes on to explain that he is suffering from allergies. The next 9 turns (lines 9-22) are taken by Greg and

Frank, who expand the discussion of allergies to include making plans to go shopping and to go outside. Note that next speaker selection is achieved implicitly through context (Sacks et al., 1974): Greg and Frank are engaged in a series of question-answer sequences about allergies and shopping, which makes them the only relevant speakers, to the exclusion of Tianyu, who self-selects as next speaker in line 23. Her self-selection as next speaker is made possible only after Frank ends his turn (line 22) and looks down at his watch, which opens up the interactional space for a third party to take the floor. It is worth noting here that Tianyu's self-selection as next speaker comes at a point in the interaction where the discussion of allergies appears to be closing, and she uses this opportunity to proffer a new topic that is relevant to her speakership: the end of the semester and her need to finish her school work.

Excerpt 1

1	Greg	¡Hola, mis amigos!
		Hi friends!
2	Frank	¡Hola!
		Hi!
3	Tianyu	¡Hola! ((saluda a Frank con la mano))
		Hi! ((waves at Frank))
4	Greg	¿Cómo estáis?
		How are you guys?
5	Tianyu	Bien, ¿y tú? ((se dirige a Greg))
		Good. How about yourself? ((turns to Greg))
6	Greg	Tengo [frío.
		I'm[cold.
7	Frank	[Más o menos. ((mueve la cabeza hacia los lados))
		[So, so. ((shakes his head))

8	Greg	¿Por qué?
		Why?
9	Frank	Em + es mucho polen outside + fuera.
		Um + there is a lot of polen + outside.
10	Greg	Yo también + más o menos? + porque necesito
		Me too + so, so? + because I need
11		nueva ropa por la primavera.
		new clothes for the Spring.
12	Frank	Ah, sí, sí ((asiente con la cabeza)), em +,
		Oh, yeah ((nods)), um +,
13		¿quiero comprar?
		do you want to go shopping?
14	Greg	Em + ¿en el centro de comercial?
		Um + at the mall?
15	Frank	Em pero yo tengo alergias y no + quiero + ir afuera.
		Um but I have allergies and I don't + want to + go outside.
16	Greg	¿Por qué no?
		Why not?
17	Frank	Tengo ALERGIAS + mucho ((tose)).
		I have ALLERGIES + I a lot ((coughs).
18	Greg	Lo siento + em + caminar en el afuera es
		I'm sorry + um + walking outside is

19		mucho bueno por tu corazón y su + em alma.
		very good for your heart and your + um soul.
20	Frank	Pero yo quiero + em comprar en el Internet ((risas))
		But I want + um to buy things online ((laughs))
21		+ muy fácil ((levanta el dedo pulgar))
		+ very easy ((thumbs up))
22		((mira su reloj))
		((looks down at his watch))
23	Tianyu	Pero + em + es la final del semestre es + em +
		But + um + the end of the semester + um + is
24		es próximo, ++ em así necesito com com completar
		close, ++ um so I need to fin fin finish
25		los trabajos de universidad.
		my papers for school.

The remainder of the scenario unfolded in a similar way: Frank and Greg took 16 of the next 19 turns. Tianyu's three turns were all self-selections at transition points. In other words, neither Frank nor Fred allocated any turns to her through explicit or implicit means, which was noted during the feedback stage (excerpt 2). In this exchange, the teacher asks the class for constructive feedback on ways to improve the scenario. In response, Sophie notes that Tianyu seemed to be left out of the scenario performance (lines 4-6). The teacher confirms the observation (lines 7-8) and asks Tianyu to account for it (line 9). The reason, according to Tianyu, was that she was unable to "find an opportunity to join in the conversation" (line 10). Her response then prompts the teacher to solicit recommendations from the class on how to "overcome" (line 13) the problem. Jane offers a general strategy ("ask her what she thinks"; lines 14, 16), and following the teacher's prompt to provide Spanish resources for doing so (line 20), Sophie and Emmitt offer two turn allocation devices: the tag question *¿y tú?* (line 21) and the Wh-question *¿Qué creas tú?* (line 23).

Excerpt 2

1 Teacher Sophie?

2 Sophie To improve?

3 Teacher Yeah.

4 Sophie I think one of the group members seems to be left out.

5 Teacher Sorry, one of the +? ((asking for repetition))

6 Sophie Group members + Tianyu.

7 Teacher Okay, yeah. Tianyu stepped in and joined the

8 conversation a bit late.

9 Why was that? ((gesturing to Tianyu))

10 Tianyu I didn't + find an opportunity to join in the conversation

11 Teacher Yeah, and that's normal, especially since you're all

12 starting to learn a new language. That's normal.

13 It's a challenge. So what could we do to overcome that?

14 Jane You kinda can ask her what she thinks.

15 Teacher Huh?

16 Jane You can ask her what she thinks.

17 Teacher Okay! So if we + if we kinda realize that the

18 other person is not talking, we can invite

19 that person to join the conversation.

20 How would we do that + in Spanish?

21 Sophie ¿y tú?

22 Teacher Yeah, ¿y tú?

23 Emmitt ¿Qué creas tú?

24 Teacher Uh huh, so we can invite that person to join us.

In sum, turn allocation in Group 1's performance was dominated by Greg and Frank who implicitly selected each other as next speaker through by expanding a topic that was not relevant to Sophie's participation in the interaction. As a third party, Sophie had to wait for opportunities to proffer a new, relevant topic. The unequal distribution of turns was noticed by the class and discussed in the feedback stage of the DSIS, during which the strategy of asking for a currently nonspeaking party's opinion, as well as two Spanish turn allocation devices, were brought up as resources for distributing turns in a more equal way.

4.2. Appropriation of a turn allocation device

Excerpt 3 displays the opening of Group 2's scenario performance (Sophie, Emmitt, and Matt). The reader will recall that two of the group members, Sophie and Emmitt, had offered suggestions for explicitly allocating next turns to nonspeaking parties in the previous feedback discussion. In the excerpt 3, we observe the deployment of the turn allocation device *¿y tú?*.

Excerpt 3

1 Sophie ¡Hola! ((saluda al grupo con la mano))

Hi! ((waves at group))

2 Emmitt ¡Hola!

Hi!

3 Matt ¡Hola! [¿Cómo estás?

Hello! [How are you?

4 Emmitt [¿Cómo estás?

[How are you?

5 Sophie No muy bien + tengo alergias primaverales.

So, so + I have Spring allergies.

6 Matt Mm.

Mm.

7 Emmitt Sí.

Oh.

8 Sophie ¿tú? ((se dirige a Emmitt))

and you?((turns to Emmit))

9 Emmitt Em + mi no tengo alergias

Um + I don´t have any allergies

10 ((mueve los hombros hacia arriba)),

((shrugs his shoulders)),

11 pero ++ [em,

but ++ [um,

12 Matt [Mm, ¿por qué?

[Mm, why?

13 Emmitt ((mueve la cabeza)) no, no ((risas))

((shakes his head)) no, no ((laughs))

14 Sophie ¿y tú? ((se dirige a Matt))

and you?((turns to Matt))

15 Matt Em + regular + I mean + necesito em +

Um + so, so + I mean + I need to um +

16 comprar ropa pero la primavera.

buy clothes for the Spring.

17 Sophie Ah, ven a ver mi casa y comprar de Internet.

Oh, then come over to my place and we can shop online.

18 Matt Pero es difícil + comprar ropa al internet por em ++

But it's difficult to + shop for clothes online cause um ++

19 ((finge que se prueba ropa)) porque soy muy em

((pretends to be trying on clothes)) cause I'm very um

20 alto y etc.? ((risas))

tall and so on? ((laughs))

21 ¿y tú? ((se dirige a Emmitt))

and you?((talks to Emmit))

22 Emmitt [Em +

[Um +

23 Matt [¿es posible em ir de compras?

[can you um go shopping?

24 Emmitt ¿ir de compras? Em + no tengo em + tiempas or

go shopping? Um + I don't have um + tima or

25 tiempo a.+ yo tengo + muchas examen y enseyas? ensayos,

time to + I have + lots of tests and essa? essays,

Following an initial greeting sequence (lines 1-4), Sophie opens the topic of spring allergies (line 5). Note that while Matt does respond (line 6), it is Emmitt who picks up the topic with a confirmation in (line 7) and to whom Sophie allocates a next turn

(line 8), which Emmitt takes up in lines 9-11. Although Matt again attempts to take the floor (line 12), Emmitt continues his turn in line 13. Upon the completion of his turn, which implicitly allocates the next turn to Sophie, Sophie turns to Matt and selects him as the next speaker, using the allocation device *¿y tú?* (line 14). In other words, she offers Matt an opportunity to join the conversation, just as she and others had discussed in the feedback stage of the previous group's DSIS task. Interestingly, Matt also appropriates the *¿y tú?* turn allocation device in line 21, where he selects Emmitt as the next speaker after a multi-turn exchange with Sophie.

During the feedback stage, several students and the teacher noted how Emmitt, Matt, and Sophie allocated turns to each other in a relatively equal way through the use of questions, including the *¿y tú?* tag. In response, the teacher asked the class for other expressions that could be used for the same function (excerpt 4). Note that Sophie offers *¿Qué crees?* as a possible turn allocation device (line 4), and the teacher in turn offers several additional resources in Spanish (lines 7-9), thus expanding the potential repertoire on which subsequent groups can draw in their DSISs.

Excerpt 4

1 Teacher Uh huh, okay, so what are other + other expressions

2 we could use instead of ¿y tú?¿y tú? + What else could

3 we say? +

4 Sophie ¿Qué crees?

5 Teacher Huh?

6 Sophie ¿Qué crees? What do you think?

7 Teacher Uh huh ¿Qué crees? ((nods)) ¿qué opinas? ¿qué piensas?

8 okay? o ¿te gusta? ¿por qué? This group, any other

9 observations?

4.3. Expansion of turn allocation repertoires

We turn now to Group 3's scenario performance. The opening is shown in Excerpt 5, where Elizabet initiates a greeting sequence, which is completed by Adam and Elijah,

respectively (lines 1-3). In line 4, then, Adam expands the greeting to ask Elijah how he is, and following Elijah's response (line 5), turns to Elizabeth to allocate the next turn to her, using the *¿y tú?* tag (line 7). What is interesting about this exchange is that Adam has expanded a typical two-turn question-answer adjacency pair into a four-turn multiparty exchange, thus creating an equal distribution of turns-at-talk (i.e., all parties are included). We note that this creates a context in which Elijah can in turn self-select to begin speaking (line 15), and who subsequently continues to engage in multiparty interaction in line 20 where he explicitly addresses both Adam and Elizabeth.

Excerpt 5

1	Elizabeth	¡Hola, amigos!
		Hi friends!
2	Adam	¡Hola, amigos!
		Hi friends!
3	Elijah	¡Hola!
		Hi!
4	Adam	¿Cómo estás?
		How are you?
5	Elijah	Así, así ((sacude la mano derecha)) em + mi + or
		So, so ((shakes his right hand)) um + me + or
6		yo tengo + ale alergias.
		I have + ale allergies.
7	Adam	Ah, sí, ¿y tú? ((se dirige a Elizabeth))
		Oh, I see. And you? ((turns to Elizabeth))
8	Elizabeth	Em + me gusta el tiem tiempos? em +
		Um + I like the wea weather? um +

9 yo quiero ir de compras?

I want to go shopping?

10 Adam Ah, sí ((mira a la ventana)) + es muy fresco,

Oh, yeah ((looks at the the window)) + it's nice

11 es primavera, pero em + no em + no puedo ir de compras

it's Spring, but um + I can't um + can't go shopping

12 porque tengo tarea y mucho em + trabaje.

because I have a lot of homework and a lot of um + work.

13 Elizabeth Oh, lo siento ((risas)).

Oh, I'm sorry ((laughs)).

14 Adam Sí,

Yeah,

15 Elijah ¿Qué clases?

For which classes?

16 Adam Em hay un examen de español esta semana +, so,

Um we have a Spanish test this week +, so,

17 Elijah [Sí,

[Yeah,

18 Elizabeth [Sí,

[Yeah,

19 Elizabeth [Em,

[Um,

20 Elijah [Puedos em puedes vamos a mi casa y em tú trabajas

[We um we can go to my place and um you can work

21 ((señala a Adam)) y tú comprar ((señala a Elizabeth)) y,

((points to Adam)) and you shop online ((points to Elizabeth)) and,

22 Elizabeth [Em,

[Um,

23 Elijah [no vamos afuera?

[we don't go outside?

24 Adam Em + pero + la biblioteca tiene los cosas que

Um + but + the library has everything I

25 ayúdame em + para estudiar,

need um + to study,

26 Em hay computadores para ir de compras, de hecho.

Um there are even computers to shop from, in fact.

A final example is given in excerpt 6. Here, the interaction primarily involves Adam and Elizabeth at the outset (lines 44-49). In line 49, Elizabeth makes a suggestion to study now and go shopping at the weekend, which is directed at Adam. However, rather than responding immediately to Elizabeth, which is Adam's right having had the turn allocated to him, he turns to Elijah and solicits his opinion with the *¿qué + crees?* turn allocation device (line 50). In other words, Adam has continued to use the expanded range of turn allocation devices that emerged from the series of DSISs in order to facilitate a relatively equal distribution of turns in this multiparty interaction.

Excerpt 6

44 Adam Por el fin de semana + ir de compras, pero estudia +

On the weekend + we go shopping, but we stud +

45 estudiamos para el [examen,

study for [the test,

46 Elizabeth [((asiente)) bien.

[(nods)) good.

47 Elijah [((asiente))

[((nods))

48 Adam esta semana?

this week?

49 Elizabeth em, sí, em + vamos en fin de semana + ahora estudiamos?

um, yeah, um + let's go shopping on the weekend + and study now?

50 Adam ¿qué + crees? ((se dirige a Elijah)

what do you think? ((turns to Elijah))

51 Elijah Em, lo siento, repito tu pre, +

Um, I'm sorry, repeat your que, +

52 Adam ¿Por la fin de semana + te quieres ir de compras y

Do you + want to go shopping over the weekend

53 por esta semana estudio? estudiamos?

and stud + study over the week?

54 Elijah Sí.

Okay.

55 Elizabeth Sí.

Okay.

5. Discussion

In this article, we have reported findings of an initial attempt at enhancing the development of interactional repertoires (Hall, 2018) through DSISs, with specific focus on turn allocation practices. In our approach to implementing DSISs into the classroom, feedback discussions were sandwiched between performances in an attempt to make feedback from the teacher as well as from other students available to subsequent groups. As we have shown, turn allocation in the first group's performance emerged as a focus of their feedback discussion because the class noticed that one of the group members had been left out of the interaction. Subsequent groups were in turn able to make use of and expand on the proposed practices for allocating turns in a more equal way (i.e., including all parties) in their performances, which we see as evidence of development. In what follows, we discuss our findings and their implications for research and pedagogy as well as some of the limitations of and future directions for our work on DSISs.

5.1. Mediating the expansion of interactional repertoires

As noted, our study aligns with and extends recent work examining pedagogical arrangements aiming to expand L2 learners' interactional repertoires (e.g., Barraja-Rohan, 2011; Kunitz & Yeh, 2019; Lilja & Piirainen-Marsh, 2019; Waring, 2019). We have focused specifically on the way in which DSIS tasks may create space for reflecting on interactional phenomena, specifically turn allocation, and raising learners' awareness of the resources available for use. As we illustrated in the analysis, focus on turn allocation practices emerged first during the feedback session on Group 1's scenario performance when a student audience member commented that one of the three group members seemed to be left out. The teacher in turn led the class in a brief discussion of strategies for distributing turns more equally in multiparty interactions. These strategies were picked up and further expanded in subsequent group performances and feedback sessions.

The major theme that emerges in this analysis is the role of the feedback session as a form of mediation (van Compernolle, 2018) leading to the expansion of learners' interactional repertoires. By mediation, we refer to Vygotsky's (1978) observation that development (i.e., the internalization of cultural tools) is supported by more experienced or competent people in one's environment who can direct and guide one's use of relevant tools. In L2 educational settings, this form of human mediation occurs in interactions in which a teacher, and possibly other learners (Donato, 1994), assist a learner in becoming aware of and using a relevant L2 resource. In other words, distinct form concepts like scaffolding and assisted performance, the notion of human

mediation focuses on the guided appropriation of a recontextualizable resource, not simply help with performance on a specific task (van Compernolle, 2015).

In the present study, mediation helped to raise learners' awareness of turn allocation as something to pay attention to during interaction as well as their awareness of some common turn allocation devices in Spanish, which they were able to put into practice in subsequent scenario performances. We believe that the proximity of mediation to performances was especially important. Since feedback sessions immediately followed performances, the class was able to reflect on and evaluate the performance while it was still fresh in their minds. In turn, the next group was able to perform their scenario shortly after the feedback session of the prior group, giving them an immediate opportunity to put into practice the suggestions that had been offered.

While our study suggests that the immediacy of mediation is an important dimension of DSIS tasks, it remains an empirical question whether, or to what extent, some delay between feedback and a next performance may impact on its effectiveness. The class we worked with was relatively small, and it was possible to have all groups perform and receive feedback in the same 50-minute class meeting. However, we recognize that such arrangements may not always be feasible, especially at larger institutions where class sizes often reach 20 students or more. In such cases, it may not be possible for all groups to perform on the same day. Teachers may therefore need to carry out DSIS tasks over multiple class meetings (i.e., all students perform over 2-3 class meetings), or alternatively to allow only a subset of students to perform each scenario. The latter suggestion would involve 2-3 small groups performing each scenario in a single class meeting, while other students would act as audience members and provide feedback. Ideally, DSISs would be performed multiple times during a term so that all students have the opportunity to perform at least one scenario. Future research would do well to examine these various arrangements of scenario performances and feedback sessions to determine the extent to which DSISs are feasible in larger classes or in courses.

5.2. The potential contribution of task repetition

It follows from the discussion of mediation in the previous section that task repetition may be one of the key strengths of our approach to using DSISs. Indeed, research in task-based language teaching supports the idea that task repetition (Bygate, 2018), or multiple iterations of a task (Larsen-Freeman, 2018), can benefit learners' performances over time. The contribution of task repetition in this research is typically explained in terms of reducing the cognitive burden of learners as they become more

familiar with the task such that they become more capable of producing increasingly complex, accurate, and/or fluent speech.

Our study is somewhat different from the cognitivist approaches to task repetition noted above in two important ways. The first difference is that the scenario performances were repeated by different groups of learners rather than by the same individuals. The second difference is that a debriefing phase in which feedback was provided was inserted between each scenario performance. Together, these two differences compel us to take a somewhat different stance toward task repetition than in previous literature, one that focuses not on the individual but on the class acting as a collective unit. Indeed, such a view aligns with Vygotskian theory and the notion of a group zone of proximal development (ZPD) that may emerge through collectivization (Petrovsky, 1985; Poehner, 2009). In this view, a group is not simply a context in which individuals act, but a collective in which individuals act "for others as for oneself" (Petrovsky, p. 191) in pursuit of the group's collective development.

Taking a group-as-collective perspective, we can see how the repetition of the scenarios by different groups and the group feedback sessions in between each performance may function as a collectivizing activity. For example, while only one group performs at a time, other members of the class are tasked with watching and evaluating the performance so that they can provide constructive feedback following the performance. This feedback session, as we have seen, serves not only to evaluate the group who has just performed, but to set the stage for the next group's iteration of the task that may incorporate the feedback and build on the prior performance. In a sense, the performances themselves are less about the individual students who happen to be performing and more about putting into practice the interactional repertoire the class is developing as a collective. Therefore, DSIS repetitions push the group's development forward, and this has the potential to benefit each individual even if they may not be current performers or direct recipients of feedback.

It should be noted that while we believe the group-as-collective perspective is useful for describing our approach to DSIS tasks, we acknowledge that our data do not allow us to assess the develop of each individual within the group. In other words, our data suggest that collectivization processes can have a beneficial impact on individuals, but we are unable to demonstrate that every member of the class learned or developed as a result of the DSIS tasks. As we continue working through our data set, we are certainly considering whether there is evidence that other students who happened not to speak during the tasks analyzed in this article expanded their interactional repertoires of turn-allocation practices. Future research should also consider developing methods for linking group and individual development over time, as discussed in the next subsection.

5.3. Tracking development over time

Our study has focused on what we consider to be the microgenesis of an expansion of students' interactional repertoires during a single class meeting involving three iterations of a DSIS. As noted, we have suggested the central role of the feedback session as mediation and the contribution of task repetition within the collective activity of the class. However, it is not clear whether, and to what extent, the potential developmental processes documented in the analysis led to growth over a longer time frame. Indeed, we would hope that DSIS tasks support learners in expanding their interactional repertoires in ways that transcend the tasks themselves so that learners become increasingly able to recontextualize the new resources they have appropriated over time. In future, we see tracking development over time as a two-part undertaking.

The first would involve tracking learners' interactional repertoires across multiple DSIS task sessions (e.g., spaced several weeks apart). Doing so would provide researchers as well as teachers specific data points for documenting students' use of relevant interactive practices as well as the opportunity to push learners by introducing more complex situations. Taking turn allocation practices as an example, we might consider designing scenarios in which next speaker selection is more complicated than a multiparty interaction among friends. For instance, introducing a unequal power dynamic in which power relations could push students to consider the ways in which power relations may be enacted through the allocation of turns as well as how and when one might self-select as next speaker as a means of challenging such a power dynamic. In other words, the point of subsequent scenarios would not be to continue to practice and expand turn allocation devices but to critically reflect on the sociopragmatic meaning and significance of turn taking and next speaker selection.

The second approach to tracking development over time would involve expanding the evidential basis beyond DSIS tasks. In classroom settings, this would primarily involve documenting students' deployment of newly appropriated interactive practices in other forms of classroom interaction. For example, the expansion of turn allocation repertoires could lead to a diversification of classroom interaction patterns in which students move out of the ubiquitous practice of Initiation-Response-Feedback (Lee, 2007; Sert, 2015; Waring, 2009). This is important because recent research has suggested that when students self-select to expand or proffer new topics and allocate turns to one another rather than always responding directly to the teacher may lead to a richer environment for learning (Adams & van Compernolle, 2019; Dolce, 2019; Dolce & van Compernolle, 2019; Waring, 2009). In other words, not only would we be interested in documenting the expansion of learners' interactional repertoires over time but also in documenting the consequences of learners' expanded repertoires for conducting the institutional business of teaching and learning in the classroom.

6. Conclusion

In this article, we have reported on the implementation of dynamic strategic interaction scenario tasks in an elementary-level US university Spanish classroom. We focused specifically on the development of turn allocation resources as part of learners' interactional repertoires (Hall, 2018). Our data suggest that the multiple iterations of the DSISs, including feedback sessions after each performance, mediated students' appropriation of relevant turn allocation practices, which they were in turn able to use in subsequent iterations of the scenario. We have also suggested that task repetition, from the perspective of the group-as-collective (Petrovsky, 1986; Poehner, 2009), was central to the microgenetic processes we observed. We are hopeful that our study can serve as a starting point for future research into the use of DSIS tasks for organizing pedagogical arrangements for the development of learners' interactional repertoires.

7. References

American Council on the Teaching of Foreign Languages [ACTFL]. (2012). *ACTFL proficiency guidelines 2012*. Fairfax, VA: American Council on the Teaching of Foreign Languages. Available at https://www.actfl.org/publications/guidelines-and-manuals/actfl-proficiency-guidelines-2012

Adams, A., & van Compernolle, R. A. (2019). *Beyond lecture-advancing responses: Knowledge imbalances and opportunities for learning in an advanced Hispanic Studies classroom*. Unpublished manuscript.

Bachman, L. F. (1990). *Fundamental considerations in language testing.* Oxford: Oxford University Press.

Barraja-Rohan, A. M. (2011). Using conversation analysis in the second language classroom to teach interactional competence. *Language Teaching Research*, *15*(4), 479-507.

Bygate, M. (Ed.). (2018). *Learning language through task repetition*. Amsterdam: John Benjamins Publishing Company.

Canale, M. & Swain, M. (1980). Theoretical bases of communicative approaches to second language teaching and testing. *Applied Linguistics, 1,* 1-47.

Celce-Murcia, M. (2007). Rethinking the role of communicative competence in language teaching. In E. Alcón Soler & M.P. Safont Jordà (Eds.), *Intercultural language use and language learning* (pp. 41–57). Berlin: Springer.

Celce-Murcia, M., Dornyei, Z., & Thurrell, S. (1995). Communicative competence: A pedagogically motivated model with content specification. *Issues in Applied Linguistics, 6,* 5-35.

Di Pietro, R. J. (1987). *Strategic interaction: Learning languages through scenarios.* Cambridge: Cambridge University Press.

Dolce, F. M. (2019). *Expanding discourse options in the foreign language classroom: Teaching turn-taking norms to enhance interactional competence in an advanced Mandarin class.* PhD thesis, Carnegie Mellon University. Pittsburgh, PA.

Dolce, F. M., & van Compernolle, R. A. (in press). Topic management and student initiation in an advanced Chinese-as-a-foreign-language classroom. *Classroom Discourse*.

Hall, J. K. (1993). The role of oral practices in the accomplishment of our everyday lives: The sociocultural dimension of interaction with implications for the learning of another language. *Applied Linguistics, 14,* 145-167.

Hall, J. K. (1995). "Aw, man, where we goin?" Classroom interaction and the development of L2 interactional competence. *Issues in Applied Linguistics, 6,* 37–62.

Hall, J. K. (2018). From interactional competence to interactional repertoires: Reconceptualizing the goal of L2 learning, *Classroom Discourse, 9,* 25-39.

Hall, J. K., Hellermann, J., & Pekarak Doehler, S. (Eds.). (2011). *L2 interactional competence and development*. Clevedon: Multilingual Matters

He, A. W., & Young, R. (1998). Language proficiency interviews: A discourse approach. In R. Young & A. W. He (Eds.), *Talking and testing: Discourse approaches to the assessment of oral proficiency* (pp. 1-24). Amsterdam: John Benjamins.

Kramsch, C. (1986). From language proficiency to interactional competence. *Modern Language Journal, 70,* 366-372.

Kunitz, S., & Ye, M. (2019). Instructed L2 interactional competence in the first year. In M. R. Salaberry & S. Kunitz (Eds.), *Teaching and testing L2 interactional competence: bridging theory and practice* (pp. 228-259). New York: Routledge.

Larsen-Freeman, D. (2018). Task repetition or task iteration? It does make a difference. In M. Bygate (Ed.), *Learning language through task repetition* (pp. 311-330). Amsterdam: John Benjamins.

Lee, Y. A. (2007). Third turn position in teacher talk: Contingency and the work of teaching. *Journal of pragmatics*, *39*(6), 1204-1230.

Lilja, N., & Piirainen-Marsh, A. (2019). Making sense of interactional trouble through mobile-supported sharing activities. In M. R. Salaberry & S. Kunitz (Eds.), *Teaching and testing L2 interactional competence: bridging theory and practice* (pp. 260-288). New York: Routledge.

Petrovsky, A. V. (1985). *Studies in psychology: The collective and the individual.* Moscow: Progress Publishers.

Poehner, M. E. (2008). *Dynamic assessment: A Vygotskian approach to understanding and promoting second language development*. Berlin: Springer Publishing.

Poehner, M. E. (2009). Group dynamic assessment: Mediation for the L2 classroom. *TESOL Quarterly*, *43*, 471–491.

Sacks, H., Schegloff, E. A., & Jefferson, G. (1974). A simplest systematics for the organization of turn-taking for conversation. *Language, 50*, 696-735.

Salaberry, R., & Kunitz, S. (Eds.). (2019). *Teaching and testing L2 interactional competence: Bridging theory and practice.* New York: Routledge.

Schegloff, E. A. (2007). *Sequence organization in interaction: A primer in conversation analysis I* (Vol. 1). Cambridge, UK: Cambridge University Press.

Sert, O. (2015). *Social interaction and L2 classroom discourse*. Edinburgh: Edinburgh University Press.

Sidnell, J. (2010). *Conversation analysis: An introduction*. West Sussex, UK: Wiley-Blackwell

van Compernolle, R. A. (2014a). *Sociocultural theory and L2 instructional pragmatics*. Bristol: Multilingual Matters.

van Compernolle, R. A. (2014b). Profiling second language sociolinguistic development through dynamically administered strategic interaction scenarios. *Language and Communication, 37*, 86-99.

van Compernolle, R. A. (2015). *Interaction and second language development: A Vygotskian perspective*. Amsterdam/Philadelphia: John Benjamins.

van Compernolle, R. A. (2018). Focus on meaning and form: A Vygotskian perspective on task and pragmatic development in dynamic strategic interaction scenarios. In M. Ahmadian & M. P. G. Mayo (Eds.), *Recent trends in task-based language learning and teaching* (pp. 79-97). Berlin: Mouton De Gruyter.

Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes*. Cambridge, MA: Harvard University Press.

Waring, H. Z. (2009). Moving out of IRF (Initiation□Response□Feedback): A single case analysis. *Language Learning*, *59*, 796-824.

Waring, H. Z. (2018). Teaching L2 interactional competence: Problems and possibilities. *Classroom Discourse, 9*, 57-67.

Waring, H. Z. (2019). Developing interactional competence with limited linguistic resources. In M. R. Salaberry & S. Kunitz (Eds.), *Teaching and testing L2 interactional competence: Bridging theory and practice* (pp. 215-227). New York: Routledge.

Young, R. F. (2008). *Language and interaction: An advanced resource book*. New York: Routledge.

Appendices

Appendix A. Activity 1

NB: A list of expressions had already been distributed for the first scenario. Subsequent DSIS preparations built on the linguistic and interactional resources.

Do some research online and add 3 new expressions to the list. For each expression, consider the following questions: What category does it belong to? What is its function? Is it formal/informal? Is it used in a specific Spanish-speaking country?

Seeking help/ repetition	Showing courtesy	Expressing doubt	Agreeing/ giving approval	Disagreeing	Making suggestions

Appendix B. Scenario description

NB: Spring A, B, and C refer to the roles for individual students. Each student received only one role description for each DSIS. We have simply compiled them here so the reader can see what all the roles were for this scenario. DSISs were adapted or inspired by scenarios compiled by Matthews (n.d.), available here: http://faculty.weber.edu/tmathews/ScenarioCollection.pdf

Topics: Leisure plans, recreation, shopping

Roles: Friends

Notions/Functions: Convincing, describing

SPRING A: Spring is finally here. You wish you could be outdoors and go on adventures, but you can't because of your Spring allergies. Invite your friends over to do something fun at your place.

SPRING B: Spring is finally here. You wish you could be outdoors and go on adventures, but you don't have time for that during the end of the semester. You have schoolwork to do and important deadlines coming up. Tell your friends what you're working on and convince them to get together to study/work/go to the library.

SPRING C: Spring is here, and you couldn't be more excited about the good weather. You can't wait to be outdoors and go on adventures, but you don't have good clothes for that. You really want to go to the shopping mall to buy sports clothes. Tell your friends about all the activities you want to do, the stores you want to go to, and convince them to come along.

Appendix C. Activity 2.

Come up with a plan and useful language for convincing your friends in the scenario to do what you want to do. Think about strategies or ways to make suggestions, offer alternatives, agree/disagree with other plans, etc. You will have a chance on Tuesday to discuss and share ideas with other classmates who are going to play the same role.

Useful expressions	Useful vocabulary	Useful grammar	Main arguments

Appendix D. Transcription conventions (adapted from van Compernolle, 2014)

+	short pause
++	long pause
.	full stop marks falling intonation
,	slightly rising intonation
?	raised intonation (not necessarily a question)

((comment)) double parentheses contain transcriber's comments or descriptions

underline underlining indicates stress through pitch or amplitude

[onset of overlapping speech

CAPITALS capital letters indicate markedly loud speech