

# The Use of Models in Collaborative Writing

---

Searching for Long-lasting Effects

Trabajo de Fin de Máster

Oskia Santesteban Marturet (~~X3M73650~~)

Supervisor: Izaskun Villarreal Olaizola

Máster Universitario en Profesorado de Educación Secundaria

Universidad Pública de Navarra (UPNA)

2016/2017

## **Abstract**

Noticing is thought to be essential for language learning (Schmidt, 1990). Various attempts have been made to promote interventions that facilitate noticing such as collaborative work or the use of models as corrective feedback. Nonetheless, the extent to which collaborative writing or the use of models affects noticing and their potential long-term effects is still under-research. It is against this backdrop where the current study must be considered. The study included 26 EFL early adolescents in a secondary school context. It was designed in three main stages: (i) the pre-test: an individual composition; (ii) the treatment: first, students had to write in pairs while their interactions were being recorded; second, they had to revise the model letter in pairs and finally, rewrite in pairs, and (iii) the delayed post-test: a second individual composition. Students' writings were analysed as for their development of the content, structure, grammatical and mechanic accuracy and vocabulary use; and their oral interactions were scrutinized for language related episodes (LREs); in addition, their final individual compositions were revised for signs of incorporations from the models. Results suggest that students benefited from collaborating and the use of models, although it seems that these advantages could not be maintained in the long-run. It is argued that memory, age and proficiency might have affected retention. Pedagogical implications are drawn and a call for further research is made as sample size and proficiency pairings might have affected the results obtained.

Key words: EFL, collaborative writing, models, effects, noticing

## INDEX

1. Introduction .....	1
2. Literature review .....	2
3. Research questions .....	7
4. Methodology .....	9
1.1 Participants.....	9
1.2 Instruments.....	10
1.3 Data collection methods.....	10
1.4 Procedure .....	11
1.5 Analysis of the data.....	12
5. Results .....	16
5.1 Research Question 1.1 .....	16
5.2 Research question 1.2 .....	18
5.3 Research Question 2.1 .....	22
5.4 Research Question 2.2 .....	23
6. Discussion .....	27
6.1 Research Question 1.1 and 1.2.....	27
6.2 Research Question 2.1 and 2.2.....	29
7. Conclusion.....	31
8. Acknowledgements .....	32
9. References .....	32
10. Appendix .....	35

## 1. Introduction

There is a growing body of literature that recognizes that noticing is essential for acquiring a second/foreign language. It is said that when a second language learner is conscious of their errors because of “self-discovering”, there is a greater chance for them to learn from those errors (Schmidt, 1990). In other words, language learners need to be able to note difficulties with specific language items as a prerequisite for learning which may happen through monitoring other more able speakers (Schmidt, 1990).

Taking advantage of the benefits of noticing, and provided teachers might be overwhelmed with writing corrections (Hanaoka & Izumi, 2012), some authors (García Mayo & Loidi, in press; Martínez Esteban & Roca de Larios, 2010; Swain & Schmidt, 1998; among others) suggest some techniques to provide feedback, such as collaborative work, peer feedback or the use of models. These techniques might be useful for teachers as for extending their options for giving written feedback (Martínez Esteban & Roca de Larios, 2010).

In the last years benefits for collaboration have been widely claimed (Fernández Dobao, 2012; Mulligan & Garofalo, 2011; Storch, 1999, 2005), as it seems that collaborating students are more likely to engage on conversations related to difficulties they are having while completing the writing task (Storch, 1999). Thus, it is likely that, in the gap-noticing process, collaboration is also beneficial for the students. In fact, many researchers have proved that collaboration enriches writings, due to the metatalk generated, among other factors (Storch, 1999, p. 364). That is to say, when students work collaboratively, they must agree on a single writing and, to do so, they discuss aspects of both content and language. This pooling of knowledge seems to result in a richer final piece of writing than when they produce it individually (Storch, 1999, 2005). Thus, in that discussion process, it is likely that learners notice more language difficulties when working in pairs or small groups than when working individually.

Besides, with the aim of assessing the effect of noticing, some authors (Coyle & Roca de Larios, 2014; Hanaoka, 2006, 2007; Martínez Esteban & Roca de Larios, 2010; Qi & Lapkin, 2001; Yang & Zhang, 2010) have explored how students of English as a foreign language use their own cognitive strategies for overcoming difficulties they could have while writing. These studies have mainly focused on two main alternatives:

the provision of native models against which learners could compare their compositions or the reformulations, meaning written feedback by the teacher. While Qi and Lapkin (2001) concluded that both reformulations and models “may be a viable way of helping non-native writers move nearer to native-writer norms” (p. 7). Other authors such as Hanaoka (2006) have claimed that models are the best form of feedback as for noticing (Hanaoka, 2006). He states that models “provide learners with alternative expressions” (p. 185) and thus, increase the possibility of improvement. That is to say, learners can produce richer texts in all senses. As expected, learners notice some problematic aspects and, then, they incorporate the improvements into their writings.

Even if positive effects of models in the short-term seem to be evident, the long-term or individual effects under-research cannot say what the long-lasting consequences are. In other words, it is uncertain yet whether those incorporations that students produce immediately after revising the models are present in future pieces of writing. The current study seeks to address this gap analyzing the consequences of models and collaboration, whether they prompt the acquisition of new features of the language for the students that make them progress in their interlanguage.

## **2. Literature review**

Little research has been done on the use of models as corrective feedback. Models make the learner be aware of their own scarcity of language. In other words, it is worthy prompting noticing of the learner’s own gaps in their second language (L2) skills. As Schmidt (1990) claims, noticing is a very important factor in Second Language Acquisition (SLA). In other words, learners need to be conscious of both their errors and difficulties while acquiring the language in order for the process to be successful.

That is the reason why some authors (Schmidt, 1990; Swain, 1985) posit the importance of output in SLA. Students are aware of their difficulties when putting the language into practice. They cannot communicate as they are used to doing in their first language (L1), which sometimes can be frustrating. As part of the process of the acquisition of a language, when learners are aware of any difficulty, they work on avoiding that scarcity of linguistic resources by correcting them in order to make themselves understood.

As noticing has been deemed essential for successful language learning, establishing what strategies or procedures boost noticing has attracted the interest of many researchers. Some authors, such as Kang (2010), position for the note-taking strategy. He hypothesized that writing any arising problem or difficulty down in a separate piece of paper could help students' noticing. It was a three-stage study in which the comparison of the results of two groups was made, one of the groups taking notes and the other one not doing so. Kang assures that this specific strategy is beneficial for causing incorporation of new items (p. 31). According to him, note-taking triggers the use of certain cognitive processes which permit the learners to incorporate more "relevant" input for their writings than those learners who do not use that particular strategy. Along the same line, García Mayo and Loidi (in press) pointed out that "guided note-taking was particularly useful in the lower proficiency levels" (p. 11), especially in order for them to focus on aspects such as content.

In contrast, some authors have shown a less positive view of the note-taking task as they have claimed that note-taking is thought to be cognitively demanding as students are asked to multi-task: note-taking and composing; thus, there might be an increased task demand (Coyle & Roca de Larios, 2014) and as a result, notes might be incomplete and might not reflect their difficulties unconstrained. In addition, some have remarked the need to apply alternative reporting techniques such as thinking-alouds and interaction recordings (Martínez Esteban & Roca de Larios, 2010; Swain & Lapkin, 1998) to monitor students' noticing. In fact, as Swain and Lapkin (1998) claim, interaction "serves as a tool [...] for L2 learning [...]" (p. 333), and that cognitive processes are evident. Neumann (2015) stated that interaction among students prompt reflection among them, which also includes reflection on language; and, therefore, they have an impact on the quality of learners' writings (p. 100). Similarly, Hanjani and Li (2014) revealed that "peers reciprocally supported each other regardless of their proficiency level" (p. 111), so both members of the pair were highly engaged in the activity. This fact might help the final writing product.

In addition, recent research studies have described the use of the L1 among low EFL learners as a tool that allows metalanguage to happen (Azkarai & García Mayo, 2015; Villarreal & Gil Sarraeta, submitted), that is students use their L1 as a scaffold to support their talk about language items or aspects of the target language when they lack the necessary target language skills to move the discussion forward in the target

language. Thus, each language would have a different role for them. The fact that students use the target language might be one more benefit for the development of it, and therefore, for the noticing of gaps in their interlanguage. The use of less constraining reporting strategies seem, therefore, more prone to supporting such type of processes.

Moreover, Storch (2005) suggested that “students may be more receptive to peers’ suggestions and feedback comments” (p. 169). This may be to say that pairs reinforce each other, so we could relate it to the idea of finding mistakes, and solving them efficiently. Following this line, if there are two people revising the same model for the same task, it is likely that they find more aspects to improve together. In this line, effects of peer feedback have been widely researched. As Wakabayashi (2008) claims, although many other factors might have been involved, this type of feedback plays an essential role in students’ writings. There is a “high degree of acceptance of peer comments” (p. 105), which can lead to the assumption that, if collaboration lets students interact for discussing different aspects of writing, those mistakes detected by their peers would be accepted and, therefore, probably solved. Even though, Sotoudehnama and Pilehvari (2016) affirm that those students who had to revise others’ writings improved more than those who had to apply the corrections. Then, students might not necessarily improve their abilities even if they accept their peers’ comments.

Following Mulligan and Garofalo’s (2011) idea, pair working is likely to benefit this aspect, as “[students] could find mistakes more efficiently” (p. 8) when working collaboratively. Likewise, Storch (1999) found that “reconstruction in pairs had a positive effect on the task performance” (p. 368). She says that students are more likely to find possible solutions to their mistakes in writing collaboratively than doing so individually, because “collaboration seemed to have a positive effect on the participants’ ability to detect the need for an amendment” (p. 368). In other words, students notice more gaps in their interlanguage when they collaborate with a pair. In fact, according to Wigglesworth and Storch (2009), collaborative writing (CW) contributes to more accurate texts, probably as a result of “pooling [students’] resources” (p. 452) and, then, choosing the best option. Similar findings are those from Shin, Lidster Sabraw and Yeager (2016), who have also supported CW as an effective tool in terms of grammatical accuracy. Furthermore, according to them, the fact of having a pair seemed

to help in improvement more than any other external variable, such as time or available resources.

Accordingly, it was likely that pair noticing was more successful than individual perceiving and, therefore, incorporation of new features. Thus, studies compared the two, pairs and individuals behaviours towards this. That is the case of Martínez Esteban and Roca de Larios (2010), who used young EFL learners of 3<sup>rd</sup> year of Secondary Education to explore this issue. It was a three-stage study in which students had to write a story in answer to a given image, revise feedback provided with (in this case models), and rewrite their previous stories according to what they had noticed. Results showed that although both individuals and pairs centred on the same issues, pairs outstood individuals in noticing and incorporating new features in their writings (p. 163). As they claim, the fact that they were grouped in pairs, make them have more linguistic resources than individuals, so in the first stage of the study, pairs noticed less. However, when they had to compare their writings with the models, pairs noticed significantly more, even if the most noticed feature were related to ideas and expression. This goes in line with Storch (2005), in that collaborative writing leads more accurate “and linguistically complex texts” (p. 163), probably as a result of the discussion of different aspects of writing during the process. This idea can perfectly be linked to that of noticing more, or at least better, that Mulligan and Garofalo (2011) highlighted as an advantage for CW.

Another point of noticing is which would be the most favourable strategy for successful noticing. Several researchers have investigated the role of noticing in reformulation and models (Adams, 2003; Hanaoka, 2006, 2007; Qi & Lapkin, 2001; Yang & Zhang, 2010). Both strategies are similar, but at the same time some aspects of them differ. The main disparity between reformulations and models lies on the fact that, whereas the former ones take learner output as a base for feedback, the latter ones are written “independently of learner output” (Hanaoka, 2006, p. 169). In other words, reformulations are done once the student has handed in the writing final product; that is, the teacher corrects their writing via writing another version of the same writing. On the other hand, models do not take into account students’ original writing product. We could consider models as “blind feedback”.



In that line, Hanaoka (2006) researched the advantages of models over reformulations, claiming that, while reformulations only provide with solutions to mistakes in the students' writing, models present features that might not appear as problems in the pre-test paper, or that students' do not address, so students might find solutions to these non-addressed problems. Hanaoka (2006) uses the term "covert" features for that type of problems. Hanaoka and Izumi (2012) confirm that "the model text [...] provides solutions to both overt and covert problems roughly equally" (p. 342). This enlightens the use of models as more effective for students than reformulations when used as corrective feedback, because not only do learners correct the previously noticed errors as a result of having the feeling of not having completed the task (Hanaoka, 2007), but they also find more possibilities that they had not considered before.

Although promising results have been reported in the studies where models have been used to boost noticing among English as foreign/second language students, the aspects in which models can be most helpful is still under-research. Some of the few existing studies, such as Coyle and Roca de Larios (2014) or García Mayo and Loidi (in press) report that noticing mostly focuses on lexis, whereas improvement in grammatical aspects are observed when using Error Correction (EC), or explicit feedback not through the use of models (Kang, 2010). According to Hanaoka (2007), students focus their revision of models on those aspects in which they have had problems when composing. Thus, if they feel a gap in lexis in their writings, their revisions will be conducted to vocabulary rather than grammar. This does not mean that they do not have holes in grammatical aspects, but that they are not aware of them. Other authors such as Martínez Esteban and Roca de Larios (2010), though, claim that, even if noticing is lexically-oriented, new ideas are what students incorporate most from the models. They disagree with the "lack of fulfillment" factor that Hanaoka (2007) pointed out, in which students searched for solutions for the gaps they had noticed before.

Whether the issues noticed by students are incorporated in their linguistic repertoire permanently is still unclear. The few authors who have studied it have reported inconclusive results that point to a lack of long-standing effects of models as written feedback. For instance, Hanaoka (2007) stated that while students improved immediately after revision, these gains did not hold in the longer period and the

improvements were not present in a delayed test. In addition, Martínez Esteban & Roca de Larios (2010) concluded that their results could be seen as evidence of learners' use of resources, but not of acquisition. Following this line, the current study seeks to determine whether the possible gains obtained from pair model revisions have long-standing effects, or they are short-time memory improvements.

Besides, young basic proficiency EFL learners are population which has not been the focus of much research. In fact, most previous research has focused on higher proficiency learners, such as Hanaoka (2007), who used two Japanese undergraduate groups of different levels (advanced and intermediate), or Martínez Esteban and Roca de Larios (2010), who explored a 3<sup>rd</sup> year of secondary education group, being their proficiency level low-intermediate. García Mayo and Loidi (in press), Hanaoka (2006), Hanaoka and Izumi (2012) or Martínez and Roca de Larios (2010, p. 159) have concluded that learners with a higher level of the target language perform better in these types of tasks. That is, they notice more than learners with a lower level. However, Coyle and Roca de Larios (2014) in their study including young learners conclude that L2 proficiency differences do not necessarily imply a variation on participants' noticing of lexical gaps.

It is against this backdrop, where the current paper needs to be framed. It seeks to explore whether 12-13 year-old low level secondary learners writing in pairs can benefit from models as a corrective strategy and whether they incorporate their noticed items into their individual linguistic repertoire.

### **3. Research questions**

The present research study seeks to explore two main issues whether collaboration facilitates noticing and hence, they note more gaps in their writings and whether models are a successful strategy to promote noticing and therefore language learning

These are the specific research questions entertained:

- 1.1. Do students write more accurate compositions when they collaborate?
- 1.2. Do students notice their gaps in their compositions when writing collaboratively? If they do, what type of gaps do they notice more? Are they able to solve them?

If the claims made by researchers advocating for the benefits of writing collaboratively are met, then it is expected that the compositions written in pairs will be more precise and likely to find solutions to their gaps than those written individually (Shin et al. 2016; Storch, 1999, 2005; Wigglesworth and Storch, 2009). Furthermore, as for the type of gaps students notice, if Coyle and Roca de Larios' (2014), Hanaoka's (2007) and Martínez Esteban and Roca de Larios' (2010) claims are corroborated then, it is expected that students will notice more lexical gaps than other types. This could be as a result of the discussion arose in students' interaction during the process of writing.

Therefore, if as pointed out by many researchers collaboration is beneficial for students' final writing products (Shin et al., 2016; Storch, 1999, 2005, 2013) as the writings created in pairs or groups are more accurate than those created by individuals on their own. As Storch (1999) affirms, final texts are more accurate when they have been written in pairs/groups than when done it individually, maybe as a result of the discussion and choice of the optimal option for the texts. Therefore, this paper will corroborate the accuracy of those final texts, as well as assess noticing in pairs' interactions.

2.1. Do students improve their writing by using models?

2.2. Do students show evidence of improvement of retain improvements in their individual post-test? If so, what aspects do they retain most?

Following the few studies that have studied the use of models, it is expected that lexis will be the most noticed type of gaps, whereas most incorporations will be new ideas. Moreover, models will have positive effects as incorporations are concerned. As a further step in the field of noticing and the incorporation of improvements in writing, this study will enlighten whether students' remember their incorporation some time after having revising the models. Following Hanaoka (2007) students are expected to incorporate their noticed language aspects into the writings immediately following the revision process, while these incorporations will not be observable at a later stage in a delayed post-test. This makes them improving after the use of models, but incorporations would not be retained.

## 4. Methodology

### 1.1 Participants

Twenty-six students in the first year of compulsory secondary education (ESO) took part in the study. Participants were 12-13 years old at the time of the study and there were 13 females and 13 males. All the students attended a private school in Pamplona.

All of them had Spanish as their L1 and they were learning English as a Foreign Language (EFL) at school. Except for one student, no students reported to have had significant contact with English outside school. This student had spent some months living abroad in the USA. Nevertheless, although he outstood his classmates in pronunciation, he did not outscore his peers in terms of grammar proficiency nor writing skills as observed by means of an individual writing task students undertook at the outset of the investigation. This initial individual writing was used as an individual competence test.

The school in which this study was undertaken works with a methodology based on tasks and projects. Thus, students are sitting in groups of four, and apart from Mathematics, and Languages (English and French or Basque), they work with interdisciplinary tasks; that is, tasks involving several subjects. It is a student-centred method, mainly, which permits the students investigate and discover new knowledge by themselves. The teachers' role is of guidance during the process. Working collaboratively was, therefore, not novel or unnatural for them as they have been working in groups for two years.

For the experimental tasks in the study, all the members of the class had to work in pairs. The school teachers' criteria were followed in order to group the students. The main factor for the grouping was that of students' working-style behaviours as setting well-functioning pairs was deemed essential for the success of the task and the reliability of the results obtained. It was considered that this decision would narrow down the causes of the difficulties experienced by learners to linguistic difficulties setting aside other non-linguistically derived problems. On the other hand, proficiency was also taken into account, as for forming pairs whose members had different levels of English. As Shin et al. (2016) affirm, those pairs formed by one low-level students and one high-level students benefit more from the collaborative task than others.

Accordingly, in this way all of the students had the chance of improving in any manner during the process.

## 1.2 Instruments

The task students had to carry out was a letter composition task. Subjects had to write an individual informal short letter, as they are used to working with this genre. Exams on the writing skills in ESO consist of the production of informal letters; for instance, to a friend. Therefore, it was deemed important to design tasks that students felt familiar with to minimize genre related difficulties. Hand in hand with the teacher the first prompt, the pre-test, was conceived as an informal email writing to an Irish student who was coming to their school in an exchange programme (see [Appendix 1](#) for the original prompt). Additionally, students' oral interactions were recorded during the pair writing task. Dyads were recorded using a voice recorder.

The third material used was an informal email model written by a native speaker (see [Appendix 2](#)). The native speaker was a teacher of the school, who was used to the level of these students. The aim was that the participants had a model of real language to focus on. Therefore, he wrote a model email that was appropriate for their level and the course standards. The classroom teacher agreed on the suitability of the model for the target participants.

The final task was the delayed post-test, where students had to write another individual piece of writing, similar to the first one in genre and topic, but including slight differences (see [Appendix 3](#)). Whereas in the first writing they were sending an email to an Irish student, in this case they had to do so to their parents, imagining that they were in Ireland. In the end, the genre and topic were the same; they had to send an email describing both the school and their experience in both cases.

## 1.3 Data collection methods

In contrast with authors such as Hanaoka (2006, 2007), Martínez Esteban and Roca de Larios (2010), or Coyle and Roca de Larios (2014), I did not use the note-taking process in order to have evidence of students' difficulties. According to Coyle and Roca de Larios (2014) this strategy might lead to a cognitive load not beneficial for the task. Learners might not complete their notes as a consequence of multi-task, so all their difficulties might not be successfully collected. Hence, following other authors

such as Hanjani and Li (2014), students’ voices were recorded when doing the collaborative task. Oral interaction prompts both members’ engagement in the task, and, therefore, not only may more gaps be noticed, but solutions for those gaps may be incorporated in future writings. Likewise, by recording students’ voices both the product and the process can be analyzed.

Even if the option of having a Control Group working individually was considered, it was finally discarded. As previously argued one of the aims of the study was to see whether collaboration was advantageous for the learners and thus, having individual and pair compositions of the same set of students was considered essential. Furthermore, a control group would have further reduced our already limited sample size and the findings would have lost power. Thus, only an experimental group (EG, hereafter) was created.

#### 1.4 Procedure

At the outset of the study students were given an individual writing task, the pre-test. Then, students did the treatment, which had three different stages (composing in pairs, pair revision of models and pair rewriting). After the treatment, they were given a second individual writing task, a delayed post-test.

Figure 1. Procedure of the study

WEEK 1	1st day (30’)	Pre-test (individual writing task)	
WEEK 2	2nd day (45’)	Treatment	Stage 1: Composition in pairs + recording of interactions
	3rd day (55’)		Stage 2: Revision of models
			Stage 3: Rewriting in pairs
WEEK 5 (two weeks later)	4th day (30’)	Post-test (delayed test) (individual writing task)	

The pre-test consisted of an individual piece of writing consisting in an informal email (see [Appendix 1](#)). Students had 30 minutes to complete the task. It was considered that the individual writing could enlighten the difficulties and level of proficiency that each student has. Unlike in other studies which used models as feedback (Coyle & Roca de Larios, 2014; Hanaoka 2006, 2007; Martínez & Roca de

Larios, 2010), the writing task was a text of a genre which students feel comfortable with. Taking that into account, it is likely that problems are mainly due to language difficulties and not to difficulties associated with the genre.

The treatment consisted of making them write a short letter in pairs (Day 2, 1<sup>st</sup> stage). After establishing the dyads, each pair was asked to write a joint informal email. 45 minutes were allotted to the task. While the students wrote the emails, their interactions were recorded. In Day 3, the same pairs were asked to retrieve the writing from Day 2 and they were provided with a model (see [Appendix 2](#)) as feedback to their writings in pairs (Day 3, 2<sup>nd</sup> stage). They were told that they had to carefully read it, focusing on features that caught their attention. No other type of feedback was given, as the study's concerns were whether the single fact of having a model makes them notice and reflect on their errors. 15 minutes were given for the deep reading of the model and they could not see their writing from Day 2, 1<sup>st</sup> stage; then, the model was removed. After having handing in the models, they had to rewrite the writing from Day 2, including as much incorporations from the models as they could remember. This final step had to be completed in 40 minutes

Day 4 was devoted to the delayed post-test. The test was done two weeks after Day 3 (3<sup>rd</sup> stage), in order to know whether students could take advantage of the incorporations they made and use them in their individual writings (3<sup>rd</sup> stage). The aim of this task was to see whether the previous noticing from the model (Day 3, 2<sup>nd</sup> stage) had long-term effects. It consisted of an individual writing, similar both in genre and topic to that done in Day 1 and Day 2 (see [Appendix 3](#)). This time, they had to write an informal email to their parents, as they were in an exchange in Ireland. The similarity between the prompts was established as a guarantee that students could incorporate the same features they included in Day 3, Stage 3. The task was to be fulfilled in 30 minutes.

### 1.5 Analysis of the data

To answer the first research question that dealt with whether students noticed gaps when writing collaboratively. The marks resulting from the pre-test and the Stage 1 of the treatment, together with the recording of the participants' voices were used. While the analysis of pre-test and Stage 1 would show a final writing result, the aim of these recordings was to know whether students had the metatalk Storch (1999) considers,

whether Language Related Episodes (LRE) occurred during the process of writing. Due to technological difficulties, half recordings were discarded and from the remaining ones five were selected. Students' dialogues were transcribed verbatim to allow for quantitative analysis. Following Araceli's (n.d.) rubric, the pre-test and Stage 1 were analysed for content and development (20%), organization and structure (20%), grammar, punctuation and spelling (30%) and vocabulary use (30%). Students' maximum score was 10. Content and development dealt with the inclusion of all the points in the instructions, and their explanations; while organization and structure referred to whether there were well-defined paragraphs, the inclusion of greetings and use of some linkers, mainly. Grammar, punctuation and spelling, as well as vocabulary use were punctuated according to their level of proficiency (basic), and what they had studied in class (see Appendix 4 for the full rubric). Scores were then entered into an Excel spreadsheet.

Meanwhile, the recordings were analysed as for the number of Language-related Episodes (LRE) created. Following Williams (1999) (as cited in Akarai & García Mayo, 2015) a LRE was considered defined as “[...] all interaction in which learners draw attention to form [...] in the context of meaningful communication as well as those that are set apart from such communication and simply revolve around questions of form itself” (p. 552). Each time students had any episode of these, they were tallied in an Excel spreadsheet. LREs were classified into [lexis-focused](#), [spelling-focused](#) and [grammar or form-focused](#), mainly. But there were also episodes of other nature, although they were not counted as LRE. These were episodes discussing ideas or expression (how to communicate an idea). The following examples illustrate the types of LRE considered.

(1) *Lexis LRE*

P8.1: When they come here we can... we... *enseñarte*  
[show you]... my school.

P8.2: ah, we show

P8.11: show, show.

(2) *Spelling LRE*



P8.2: I hope that you like...

P8.1: [writing]... Like...

P8.2: and come.

P8.1: [showing insecurity]

P8.2: [spelling] c-o-m-e. *Y ya.* [and that's all]

### (3) *Grammar LRE*

Rosa: I think it's going

Julián: Going, going! *Porque es su... sujeto de la oración.*

[Because it is... the subject of the sentence.]

(Azkarai and García Mayo, 2015, p. 557)

Unfortunately, no grammar-focused LREs were observed in the study.

For the second part of the investigation, Stage 3 and delayed post-test were compared. Firstly, the scores resulting in each one of the steps were contrasted in order to know about the improvement triggered by models. Secondly, the number of incorporations from the model included in Stage 3 of the treatment, and in the delayed test was analyzed. Hanaoka's (2007) and Martínez and Roca de Larios' (2010) criteria were followed. So "a feature attempted in the revision" (p. 465) counted as incorporation even if it was mistaken. That is to say, when incorporation is made, but its use is not correct. Example (4) shows an attempt of this type.

### (4) *Mistaken incorporation*

(P1): "I have bad \*notes" (marks)/

"It is \*semi-priate" (semi-private)

Once again, the incorporations were tallied in Excel and the results from the rewriting and the delayed post-test were compared to see whether incorporations were maintained individually after two weeks.

As it was done with LREs, incorporations were also classified into different types. Five types of incorporations were considered: lexical incorporations (see example (5)

below), introducing any new lexical item that they did not know before or that they had not had considered before; mechanical incorporations (see example (6)), correcting any spelling mistake they had; grammatical corrections (see example (7)), correcting any grammatical point they had misused; incorporation of ideas (see example (8)), whether students incorporated new ideas appearing in the model that they had not considered previously; incorporations of expression (see example (9)), which included correct structure of the genre, punctuation and order of sentences. The following examples show the nature of each one of the categories.

(5) *Lexical incorporation*

(P5): “It has a part for babies [...]” – “It has a nursery room [...]”

(6) *Mechanical incorporation*

(P3): “\*sporthall - sportshall” / “\*graound - ground”

(7) *Grammatical incorporation*

(P8): “[...] \*you can came [...]” – “you can come”

(P9): “\*The things that [...] is” – “The things that [...] are”

(8) *Incorporations of ideas*

(P4): “The school is in Chantrea” (location)

(P8): “[...] school in Spanish”

(9) *Incorporations of expression*

(P4): “\*Dear Irish people:” - “Dear Irish people,”  
[punctuation]

(P5): “bye, kisses” [greetings]

Summing up, to examine whether CW lead to more accurate texts, students' scores in their pieces of writing were analyzed. To explore the effect of CW on noticing, interaction recording was done, searching for LREs. Similarly, the investigation of the effects of the use of models was based both on the final scores and on the incorporations

students made in Stage 3 and the delayed post-test. The next section includes the results obtained from the pre-test, the treatment (stages 1, 2 and 3) and the post-test.

## 5. Results

In this section, the results from the collected and analyzed data will be presented. As previously stated, there were four research questions to be answered, so the section is structured as follows. The first subsection deals with the results that aim to answer the first research question block on collaboration, which is subdivided in two (RQ 1.1) whether students write more accurate compositions in CW, and (RQ 1.2) whether CW permits noticing as observed by the generated LREs. The second part, targets the second research question pair, which deals with (RQ 2.1) whether students overall improve using models and (RQ 2.2) whether the improvements are observable in the individual post-test as evidenced by the number of incorporations.

### 5.1 [Research Question 1.1](#)

First question was to confirm whether students wrote more accurate texts in collaboration. The average scores of individual writings and composition in pairs (Stage 1 of the treatment) were compared, apart from analyzing their corresponding voice recordings.

[Table 1](#) shows the differences between the marks from individual writing in the pre-test and the pair writings in Stage 1. As it is shown, data from 13 pairs were collected. Each member of the pairs has been concreted as well (P1.1/P1.2...).

Results show that seven of the pairs obtained higher scores when working collaboratively. Some of them improved remarkably, as for P4 (from 9.25/8.5 to 10), P10 (from 7/5 to 8.5), or P12 (from 6.75/8.5 to 9.5), whereas other pairs' improvements were more subtle (P5, P6, P7). However, in another 5 pairs one of the members, the member that scored lower in the individual writing improved; for instance, P2 (from 8.5/9.25 to 8.75), P3 (from 6.75/5.5 to 5.75), P6 (from 10/8.5 to 9), P8 (from 4.75/8.7 to 8.75), P9 (from 8/8.75 to 8.5). Only pair that does not meet this pattern is P3 (from 6.75/5.5 to 5.25).

Table 1. Comparison of individual and pair marks

		<b>INDIVIDUAL MARK (pre-test)</b>	<b>PAIR MARK (treatment Stage 1)</b>
<b>P1</b>	<b>P1.1</b>	6.75	6.75
	<b>P1.2</b>	5	
<b>P2</b>	<b>P2.1</b>	8.5	8.75
	<b>P2.2</b>	9.25	
<b>P3</b>	<b>P3.1</b>	6.75	5.25
	<b>P3.2</b>	5.5	
<b>P4</b>	<b>P4.1</b>	9.25	10
	<b>P4.2</b>	8.5	
<b>P5</b>	<b>P5.1</b>	8.5	9
	<b>P5.2</b>	8.75	
<b>P6</b>	<b>P6.1</b>	10	9
	<b>P6.2</b>	8.5	
<b>P7</b>	<b>P7.1</b>	8.5	8.75
	<b>P7.2</b>	8.5	
<b>P8</b>	<b>P8.1</b>	4.75	7.75
	<b>P8.2</b>	8.75	
<b>P9</b>	<b>P9.1</b>	8	8.5
	<b>P9.2</b>	8.75	
<b>P10</b>	<b>P10.1</b>	7	8.5
	<b>P10.2</b>	5	
<b>P11</b>	<b>P11.1</b>	4.75	9.5
	<b>P11.2</b>	7.75	
<b>P12</b>	<b>P12.1</b>	6.75	9.25
	<b>P12.2</b>	8.5	
<b>P13</b>	<b>P13.1</b>	5.25	8.75
	<b>P13.2</b>	7.25	

As [Table 2](#) shows, there is improvement from the pre-test to Stage 1 (from 6.62 to 8.43). We can also see improvement from the pre-test to the delayed post-test (from 6.62 to 7.19), which assures that CW leads to a progress in writing. Nevertheless, if we compare Stage 1 (8.43) and the delayed post-test (7.19), there is a clear decrease in scores. The aspect in which students generally get most points is content and development (3.69 in Stage 1; 3.35 in the delayed post-test). Nevertheless, although that category is the most improved one in the delayed post-test (from 3.07 to 3.35), it is not the same category the one in which they improve most in Stage 1, but grammar, punctuation and spelling (from 2.32 to 3.38).

Tabla 2. Improvements.

	pre-test		Stage 1		post-test	
	points	score	points	score	points	score
<b>Content and development</b>	3.07	7.68	3.69	9.23	3.35	8.38
<b>Organization and structure</b>	3.07	7.68	3.38	8.45	2.64	6.6
<b>Grammar, punctuation and spelling</b>	2.32	5.8	3.38	8.45	2.82	7.05
<b>Vocabulary use</b>	2.42	6.05	3.15	7.88	2.78	6.95
<b>AVERAGE</b>	2.72	6.62	3.4	8.43	2.90	7.19

Overall, it can be affirmed that CW leads to more grammatically accurate texts, but when they write individually in the delayed post-test, students do not produce as accurate texts as in CW. Anyways, students improved their scores in writing, both when working in pairs and when lately individually assessed.

## 5.2 [Research question 1.2](#)

Research question 1.2 aimed to know collaboration led to noticing as evidenced by the LREs attested. As noticing is concerned, collaborative writing did seem to trigger interaction among members of the pairs. [Table 3](#) shows the nature of the interactions among the pairs.

Table 3. Oral interaction between pairs

	LREs				ORGANIZATION OF IDEAS			TOTAL EPISODES
	lexis	spelling	grammar	total LREs	ideas	expression	total organization	
<b>P6</b>	3	1	0	<b>4</b>	8	0	<b>8</b>	12
<b>P7</b>	3	0	0	<b>3</b>	5	1	<b>6</b>	9
<b>P8</b>	1	2	0	<b>3</b>	2	1	<b>3</b>	6
<b>P10</b>	4	1	0	<b>5</b>	6	0	<b>6</b>	11
<b>P13</b>	0	0	0	<b>0</b>	4	0	<b>4</b>	4
<b>Total</b>	11	4	0	<b>15</b>	25	2	<b>27</b>	42

Table 3 shows the total number of LREs P6, P7, P8, P10 and P13 had in their interactions. As we can see, the majority of the pairs had more episodes related to organization of ideas than LREs. For instance, P6 had four LREs in their interactions, whereas 8 episodes of organization of ideas were present. In fact, P13 does not create any LRE while they discussed ideas 4 times.

Table 4 shows which amount of LRE ended up resolved and which did not. Recorded pairs showed a tendency to reach an agreement when confronted with some linguistic difficulty as not resolving the difficulty was the least popular option for the pairs (3/20, 20%). The results show that overall in almost half of the instances (46.66%) students correctly-solved the LREs, while in 33.33% of the instances they inappropriately resolved them (5/20). As Storch (2005) or Wakabayashi (2008) claim, dialogues during the process of writing collaboratively are important for learning, since students are thought to accept and be receptive to their peers' feedback.

Table 4. Amount of LREs solved, unsolved or wrongly solved

pairs	LREs	solved	%	unsolved	%	wrongly solved	%
P6	4	2	50	2	50	0	0
P7	3	0	0	0	0	3	100
P8	3	3	100	0	0	0	0
P10	5	2	40	1	20	2	40
P13	0	0	/	0	/	0	/
<b>TOTAL</b>	<b>15</b>	<b>7</b>	<b>46,66</b>	<b>3</b>	<b>20</b>	<b>5</b>	<b>33,33</b>

Some extracts from the recordings can be seen in the examples below. The extract shows that, as it was attested by Azkarai and García Mayo (2015), the participants use the L1 and the target language for solving the linguistic problems they encounter. Nevertheless, the recordings show that students use English for a fair amount of time which is another positive point of Collaborative Writing as “pair work provides learners with many opportunities to use their L2” (Azkarai & García Mayo, 2015, p. 560). [Example 10](#) shows us an extract from P10 in which they discuss a lexis gap.

(10) P10.1: *¿cómo se dice “taller”?* [How do you say “taller” [studio]?:]

P10.2: em... artistic room?

P10.1: *no sabemos cómo se dice taller.* [We don’t know how “taller” is said.]

P10.2: *ponemos “artistic room”. Podemos poner entre paréntesis taller.* [We’ll write “artistic room”. We can write “taller” in brackets]

In [Example 10](#), participants were discussing about a gap in vocabulary. They were not sure about how *taller* (studio) is said in English, so they decided to write it in another way (“artistic room”).

Gaps in spelling were also discussed, as we can perceive in [Example 11](#). The extract was taken from P8. In this case the focus is on mechanics, a member of the dyad spells out the word healthy to his partner. That is to say, one of the mates has a gap, and the other participant solves it. Following Storch (1999), this is beneficial to the task, as they are more likely to learn through this process.

(11) P8.2: It is very healthy. Heal-thy.

P8.1: Is like that?

P8.2: h-e-a-l-t-h-y

P8.1: ah! *Ah vale* [oh OK].

In contrast, we can see some other extracts in which one member of the pair corrects some errors in the language the other one commits. This is the case of [Example 12](#). P6 was discussing about what to write in their composition when P6.2 made a mistake: he misused the plural of the word “person”. P6.1 was very attentive to correct it.

(12) P6.2: no eh... we can put that is... is a... it's like a group of persons

P6.1: People [shouting]

P6.2: people. OK.

However, this does not mean that students could retain what their partners were saying. Some other times it seems that even if a student knows the word fails to correct his/her partner as in the next passage. [Example 13](#) illustrates an attempt of lexis-related LRE, showing how P6 ignores the word that his partner is using, and keeps on the use of an invented word (unresolved LRE). In this case, P6.2 uses this word in the whole recording.

(13) P6.1: we do a lot of subjects.

P6.2: we have a new “asignature” (subject)... that is MEDAP.

Hence, students discussed between them which were the best options of language to include in their writings. Nevertheless, as it can be seen in [Example 12](#), the fact that a partner has a higher level of proficiency in the language does not guarantee that he/she will be able to help the other member of the pair notice their gaps.

To summarize all the block, like authors such as Fernandez Dobao (2012) or Storch (1999, 2005) affirm, our results show compelling evidence for some of the participants, whereas not for others, who seem to be the ones with a higher level of



English. We can also affirm that CW generally permits better noticing, as well as solving mistakes after the revision of models.

### 5.3 [Research Question 2.1](#)

As for whether the use of models leads to an improvement in learners' writings, marks in the pre-test individual writing and the post-test individual writing were compared, as to investigate whether students take advantage of that improvement in terms of writing skills. [Table 5](#) shows this comparison. The same rubric in [Appendix 4](#) was used.

As can be seen, models have a positive effect in terms of improvement, as the marks' mean scores increase (from 7.29 to 7.89). It is true that 8/26 students did not improve their writings (P1.2, P3.1, P5.1, P5.2, P6.1, P8.2, P9.2 and P11.2). Some others (5/26) got the same mark as in the pre-test (P1.1, P2.2, P4.1, P4.2 and P12.2). Half of the students (13/26), though, got higher marks; some of the individuals even improved remarkable, as P 10.1 (from 7 to 10), P10.2 (from 5 to 7.5), P11.1 (from 4.75 to 7.75), or P13.1 (from 5.25 to 8.75).

As we have observed marks in the post-test are generally either equal or higher than those in the pre-test. Only eight people did not get a higher score. It seems, then, that noticing through the use of models might play a role in those improvements.

Table 5. Individual marks from the pre-test and post-test

		<b>Pre-test</b>	<b>Delayed post-test</b>
<b>P1</b>	<b>P1.1</b>	6.75	6.75
	<b>P1.2</b>	5	3
<b>P2</b>	<b>P2.1</b>	8.5	8.75
	<b>P2.2</b>	9.25	9.25
<b>P3</b>	<b>P3.1</b>	6.75	6.5
	<b>P3.2</b>	5.5	8.75
<b>P4</b>	<b>P4.1</b>	9.25	9.25
	<b>P4.2</b>	8.5	8.5
<b>P5</b>	<b>P5.1</b>	8.5	7.25
	<b>P5.2</b>	8.75	8.25
<b>P6</b>	<b>P6.1</b>	10	8.25
	<b>P6.2</b>	8.5	8.75
<b>P7</b>	<b>P7.1</b>	8.5	10
	<b>P7.2</b>	8.5	8.75
<b>P8</b>	<b>P8.1</b>	4.75	6.25
	<b>P8.2</b>	8.75	8.5
<b>P9</b>	<b>P9.1</b>	8	8.25
	<b>P9.2</b>	8.75	8.25
<b>P10</b>	<b>P10.1</b>	7	10
	<b>P10.2</b>	5	7.5
<b>P11</b>	<b>P11.1</b>	4.75	7.75
	<b>P11.2</b>	7.75	7.25
<b>P12</b>	<b>P12.1</b>	6.75	7.25
	<b>P12.2</b>	8.5	8.5
<b>P13</b>	<b>P13.1</b>	5.25	8.75
	<b>P13.2</b>	7.25	8
<b>MEAN</b>		7.29	7.89

#### 5.4 [Research Question 2.2](#)

Research question 2.2 explored whether students showed signs of incorporation of previously commented aspects actually. Thus, the rewriting in stage 3 in the treatment and the writing resulting from the post-test were compared. Hence, I could see whether participants incorporated the same improvements to their writings as they did in the rewriting.

Table 6. Number of incorporations per pair in stage 3

	<b>Lexis</b>	<b>Spelling</b>	<b>Grammar</b>	<b>Ideas</b>	<b>Expression</b>	<b>Total number of incorporations</b>
<b>P1</b>	1	0	0	2	0	3
<b>P2</b>	0	0	0	4	2	6
<b>P3</b>	1	6	4	0	0	11
<b>P4</b>	0	0	0	1	0	1
<b>P5</b>	1	0	0	1	1	3
<b>P6</b>	1	0	1	3	0	5
<b>P7</b>	1	0	2	1	0	4
<b>P8</b>	0	1	0	2	0	3
<b>P9</b>	2	1	0	3	2	8
<b>P10</b>	0	0	1	3	0	4
<b>P11</b>	2	0	0	3	1	6
<b>P12</b>	0	0	0	5	0	5
<b>P13</b>	0	0	0	4	0	4
<b>TOTAL</b>	9	8	8	32	6	63

Table 6 illustrates the number of incorporations that each pair included in their rewritings in stage 3, just after the revision of the model. Taking into account that compositions were of 80-100 words long, it can be stated that every pair included quite a few improvements to their writings; some writings included more than 5 improvements, which clearly outstand the rest (P2, P3, P9, P11). Surprisingly, these results are not correlated to that of the scorings; in other words, those four pairs (P2, P3, P9, P11) do not outscore the other pairs.

Table 7 shows that ideas (50.79%) were the items which students most incorporated. In fact, incorporation of ideas represented half of the incorporated aspects. Lexis, grammar and spelling represent the next most popular type of incorporations but the percentage drops down to around 14%. Expression is the type of item with the lowest proportion of incorporation, 9.52%.

Table 7 Number of classified incorporations in Stage 3

	<b>n.</b>	<b>%</b>	<b>Mean number of incorporations</b>	<b>s.d.</b>
<b>lexis</b>	9	14,29	0,69	2,334641182
<b>spelling</b>	8	12,70	0,61	2,537628901
<b>grammar</b>	8	12,70	0,61	2,282277636
<b>ideas</b>	32	50,79	2,46	8,01646657
<b>expression</b>	6	9,52	0,46	1,65748386
<b>TOTAL</b>	63	100	4,83	16,82849815

When their final writings in the delayed post-test were analyzed for traces of these incorporations; however, it seemed that participants overall showed very few incorporations in their individual tests. While 63 incorporations were detected in Stage 2, only 11 incorporations were observable in the post-test. Yet, once again ideas were among the most incorporated items, but surprisingly almost all the expression items were retained in this final writing (5 out of 6) being the most frequent one (45.4%). In contrast to this, no grammatical item was incorporated and only a single lexical or mechanical aspect which had been previously noticed has been incorporated in the final individual task (delayed post-test).

Table 8. Improvements included in the delayed test

	<b>n.</b>	<b>%</b>	<b>Mean number of incorporations</b>	<b>s.d.</b>
<b>lexis</b>	1	9,09090909	0,038461538	0,19612
<b>spelling</b>	1	9,09090909	0,038461538	0,19612
<b>grammar</b>	0	0	0	0
<b>ideas</b>	4	36,3636364	0,153846154	0,4641
<b>expression</b>	5	45,4545455	0,192307692	0,40192
<b>TOTAL</b>	11	100	0,423076923	1,25826

If results from the treatment writing are compared against the results from the delayed post-test (see [Table 7](#) and [Table 8](#)) it can clearly be seen that the total number of improvements included are less in the delayed test than in stage 3 of the treatment, 63 vs. 11, respectively. There is not a single grammatical incorporation into the students' individual writings and only 4 ideas have been retained (from the total of 32 in Stage 3). It seems that the number of incorporations drops dramatically from Stage 3 to the delayed post-test.

In order to better analyze data on incorporations, length of writings needs to be established. Table 9 shows the length of each one of the products the students handed in.

Table 9. Number of words in writings

		Pre-test	Stage 3	Delayed post-test	
<b>P1</b>	<b>P1.1</b>	128	102	120	<b>P1.1</b>
	<b>P1.2</b>	63		54	<b>P1.2</b>
<b>P2</b>	<b>P2.1</b>	109	150	130	<b>P2.1</b>
	<b>P2.2</b>	154		153	<b>P2.2</b>
<b>P3</b>	<b>P3.1</b>	70	64	80	<b>P3.1</b>
	<b>P3.2</b>	54		64	<b>P3.2</b>
<b>P4</b>	<b>P4.1</b>	120	117	96	<b>P4.1</b>
	<b>P4.2</b>	93		108	<b>P4.2</b>
<b>P5</b>	<b>P5.1</b>	113	130	99	<b>P5.1</b>
	<b>P5.2</b>	100		85	<b>P5.2</b>
<b>P6</b>	<b>P6.1</b>	116	172	108	<b>P6.1</b>
	<b>P6.2</b>	88		99	<b>P6.2</b>
<b>P7</b>	<b>P7.1</b>	135	96	91	<b>P7.1</b>
	<b>P7.2</b>	67		90	<b>P7.2</b>
<b>P8</b>	<b>P8.1</b>	79	77	90	<b>P8.1</b>
	<b>P8.2</b>	85		92	<b>P8.2</b>
<b>P9</b>	<b>P9.1</b>	90	132	89	<b>P9.1</b>
	<b>P9.2</b>	115		105	<b>P9.2</b>
<b>P10</b>	<b>P10.1</b>	109	81	94	<b>P10.1</b>
	<b>P10.2</b>	61		49	<b>P10.2</b>
<b>P11</b>	<b>P11.1</b>	52	164	84	<b>P11.1</b>
	<b>P11.2</b>	147		85	<b>P11.2</b>
<b>P12</b>	<b>P12.1</b>	80	122	66	<b>P12.1</b>
	<b>P12.2</b>	104		90	<b>P12.2</b>
<b>P13</b>	<b>P13.1</b>	88	81	122	<b>P13.1</b>
	<b>P13.2</b>	107		109	<b>P13.2</b>

As we can see, from pre-test (individual) to Stage 3 (in pairs), there are only one pair (P8) who produces a shorter writing (from 79/85 to 77). Some other five pairs (P1, P2, P7, P10 and P13) only the member who had written a shorter text in the pre-test gets to improve in length. Then, nearly most of the pairs wrote more words in collaboration (Stage 3) than individually (delayed test). Some of the pairs, though, did not get a longer writing product, as for P3, P8 and P13. The case in these last dyads are that the one writing longer texts keeps on writing long, whereas the one who writes shorter, tend

to write shorter than his peer. Surprisingly, though, most of the individuals write shorter texts in the delayed post-test, whereas only eleven individuals got longer texts (P2.1, P3.1, P3.2, P4.1, P6.2, P7.2, P8.1, P8.2, P11.1, P13.1, and P13.2). As the length of the writings lowers, the number of incorporations does so as well.

To sum up, participants did include a considerable amount of changes taken from the model in the immediate rewriting of their writings. They overall included more ideas than any other type of improvement. Notwithstanding, the delayed tests show us that those improvements are generally not retained in the long run by the individual students as demonstrated by the few instances incorporated in the individual attempts.

## **6. Discussion**

The present study was designed to determine the effect of pair noticing through models in CW. Two main research question blocks were stated, each of them having a different aim in the study. The first two research questions aimed at determining whether participants in the study wrote more accurate writings as measured through a general rubric, as well as whether they were able to notice their own gaps in collaboration through the evidence gathered in the LREs. The second set of research questions, explored the extent to which the use of models as the only type of feedback provided for their writings was beneficial, paying attention to improvement of the final product, as well as the number of solved noticed gaps remembered. In what follows the research questions raised will be discussed under the hypotheses entertained.

### **6.1 Research Question 1.1 and 1.2**

In light of the benefits of CW shown by other authors (Fernández Dobao, 2012; Storch 1999, 2005; Villarreal & Gil Sarraeta, submitted), the two first questions in this study sought to determine whether students wrote more accurate compositions and noticed their gaps when collaborating in their writing tasks. For that, the first individual writing (pre-test), the first writing in pairs (Stage 1) and the last individual writing (delayed post-test) were compared. Students' oral interactions were also recorded to explore the LREs created.

The results obtained in the study confirm findings in other studies in that CW leads to more accurate texts (Storch, 1999; Villarreal & Gil Sarraeta, submitted). There

was evidence to affirm that this strategy allow learners to pool their knowledge of the language and therefore, produce more grammatically accurate texts. Nevertheless, in contrast to Swain and Lapkin's (1998) claim that "students tended to retain the knowledge that they had constructed collaboratively" (p. 323), this accuracy seemed not to be retained, as scores in the delayed post-test decreased, being grammar one of the categories in which students showed less retention. Then, it is suggested that discussion and collaboration might not be sufficient conditions for acquisition to occur. Hence, individual noticing might be one of the required conditions for acquisition, or there might be other factors such as age or proficiency that intervene in the process. Notwithstanding this, general scores in the delayed post-test were still higher than those of the pre-test; hence, CW effects cannot be dismissed.

The current study also found that, together with getting higher accuracy scores in their writings when collaboration happened, students noticed gaps, or linguistic difficulties, in their EFL language. As Storch (1999) affirms, metatalk happened among the pairs. That could have been a signal of thinking about their gaps in output, implying the searching for solutions. In our data, some evidence of LRE was found, and, moreover, most of the gaps were correctly resolved. LRE are important as evidence of metatalk, or reflection on students' own knowledge of the language. In other words, this type of interactions implies students pool their knowledge together to overcome language difficulties they notice, the gaps. As pointed out before, Schmidt (1990) and Swain (1985) among others state that noticing is necessary to progress in language learning. Furthermore, the fact that students are able to solve many LREs positively reinforces the importance of pair work as a valid scaffold to support language learning. Therefore, in consistency with findings in Storch (2013), students better retain feedback if given by their peers in the collaborative writing setting.

Even though many of the gaps were solved, there were some others which were left unresolved, even though they were the least frequent ones. It is argued that probably this is due to a lack of communication breakdown. When the gaps students might have do not hinder comprehension, students do not interrupt the communication flow to discuss linguistic aspects. Frequently, students fill these gaps with resources extracted from a shared knowledge about mother tongue or social reality which does not necessary conform to the target-language norms. This is evidenced in [example 10](#) and

[example 13](#) where students include vocabulary items which are not part of the target language but allow the progress in their discussion.

This strategy might be partly related to the level of proficiency of the participants. According to Hanaoka (2006, 2007), those participants with a higher level of English proficiency notice more gaps than participants with lower proficiency levels. In this case, though, it seems that it is the lower proficiency students that improved the most in most of the cases. The cause of this could have been that pairs were not formed by members with the same proficiency level. That is, higher level learners had to form pair with lower level ones, so that whenever one of the members was not aware of a gap, the partner could assist him/her. Although pairing of uneven pairs have been successful according to what had been predicted, next time students might be paired according to similarity in levels of proficiency. That would give us a broader sight on how proficiency might be a variable.

## 6.2 Research Question 2.1 and 2.2

Regarding the improvement due to the use of models, writings in the delayed post-test outscored those in the pre-test. Models might have had an effect on that, as they help students notice their gaps in writing. In this case, collaboration might also have an effect, so that both aspects together might have produced an increase of this improvement.

On the question of whether students actually incorporated from the models in the delayed post-tests as well, this study confirmed that they did generally not include their previous improvements in their future writings (delayed test). Agreeing with previous studies using delayed tests (Hanaoka, 2007), our findings showed that there were fewer incorporations in the delayed test than in the immediate rewriting after the revision of models (Stage 3). However, it should be taken into consideration that, even if incorporations in the delayed post-test decreased, the average length of the writings did so as well. Therefore, the number of incorporated items might have decreased as well.

In any case, the possible interference of memory cannot be ruled out. It was predictable that when participants were asked for writing after two weeks in this case, they did not remember the improvements they had made to their previous tasks, even if the new one was a similar task. As Robinson (1995) states, noticing must be honed in



the short-term memory in order to be linked in the long-term memory. That implies that, if this practice of noticing from models had been repeated more than once, students might have remembered the improvements in Stage 3 more effectively.

In this study, participants incorporated new ideas, mainly, and not new vocabulary nor grammar. These findings are thus in line with Martínez Esteban and Roca de Larios (2010), in which participants include more incorporation on ideas and expression. The similarity with Martínez Esteban and Roca de Larios (2010) might be due to participants' age proximity (12-13 year-old teenagers in this paper, 15 year-old ones in the former). Other studies, however, have reported contradictory findings where lexical items were the most frequent incorporations (Coyle & Roca de Larios, 2014; Hanaoka, 2007). These two studies, however, deal with adult or children participants which may have influenced the results as the age of the participants might be paramount as a variable. Even if the well-known critical period for acquiring the language benefits young learners, late starters also benefit from "cognitive maturity" (Muñoz & Singleton, 2011, p. 17), as they have better test-taking skills, and better use of strategies. Likewise, attention adolescents can pay both to their manner of writing and to models is assumed to be more limited than in the adults' case. Noticing, then, might be more efficient with older learners than younger ones.

An additional factor affecting the results obtained could be the proficiency of the participants involved. Hanaoka (2007) used undergraduate advanced students whereas Martínez Esteban and Roca de Larios (2010) used low-intermediate secondary students, both of which ended up with different results regarding type and number of gaps noticed. This paper, though, investigates a classroom of Secondary Education with a low level of proficiency. Results seem to be more in line with Martínez Esteban and Roca de Larios (2010), so the proficiency variable might also have an effect on the type of gaps noticed, since more proficient learners are more used to the English grammatical structures or terms. Nevertheless, proficiency level does not necessarily have a direct relationship with the number of incorporations it has been proved in a way that, as Coyle and Roca de Larios (2014) pointed out. In this paper, the pairs that incorporate more are not necessarily the most proficient ones, or the ones that get higher scores. Then, proficiency might have an effect on the type of gaps noticed, but not on the number of incorporations.

## 7. Conclusion

This paper explores the effect collaborative writing and the use of models have in SLA. Models could be a useful tool as written feedback to students. Via a text written by a native speaker, feedback to students' writings can be provided. Concerning that they had great benefits, they could be used in teaching. Short-term effects of models are known (Coyle & Roca de Larios, 2014; Hanaoka, 2007; Martínez Esteban & Roca de Larios, 2010). However, there is not much research on the long-standing effects of models.

Firstly, it was investigated whether collaborative writing results in more accurate texts, as well as whether students noticed their gaps when they worked in collaboration. Secondly, how models lead into improvement of the writings was studied. It was contemplated the use of a delayed post-test, as it seemed of importance to know whether students included improvements some weeks after revising the models, leading to know whether models also have long-lasting effects.

One of the most significant findings to emerge from this study is that, evidence was found on the fact that CW benefits noticing, which might have resulted in more accurate final writing products. Learners had LREs in their interaction, which lead to the thought that they noticed gaps when they collaborated. Furthermore, most of the gaps appearing in their LREs were found a solution. The fact that there were members of different levels in each pair benefited them; when one of the members was not aware of a mistake, the partner could help him/her. Overall, it can be affirmed that collaboration helped pairs be aware of their gaps.

Likewise, comparing the writing final products in the immediate stage after revising models and the ones in the delayed test, the amount of incorporations from the model clearly lowers. This is to say that students do not retain the improvements in their long-term memory. One possible reason why this happened could be that the type of incorporation that students most did was new ideas. This is in unison with Martínez Esteban & Roca de Larios (2010), but not with Hanaoka (2007), who detected more incorporation in lexis. Proficiency might have had an effect on the type of noticed gaps, but in is not likely that it had any effect on the number of incorporations. At any rate, improvement in writings after the treatment cannot be rejected.

The most important limitation lies in the fact that this process has been done one time only; maybe if the same practice had been repeated, more specific benefits could have been found. Likewise, the fact that a control group was not used restricts the generalization of the findings too. Notwithstanding limitations, the findings of this study have a number of important implications for future teaching practice. The use of models as a feedback tool is beneficial in the short-run. However, as for the long-term effects are concerned, another complementing tool should be used in order for new items to be acquired in the process of SLA.

More research using controlled trials is needed to confirm how CW could affect noticing. Even if this study has found evidence on CW and models' effects on awareness, it is likely that limitations have restricted these findings. A control group working individually should be used as for comparing noticing in individuals and in pairs. It should also be specified whether the improvements in scores that have been found in this study are due to CW or to the use of models, imposing different variables or even two different studying groups. On the other hand, further research is needed on the long-standing effects of models in EFL, as it seems to be a beneficial tool for L2 writing that needs not to be ignored. If the use of models in EFL had long-lasting effects, they would be a useful written feedback strategy to be used in the classroom. Teachers should be trained in these strategies, as it seems that these techniques are likely to be effective tools as for learners' development of their interlanguage.

## **8. Acknowledgements**

I would like to express my gratitude to Dr. Izaskun Villarreal for her guidance and recommendations since I started with the idea of studying in this field. Without her valuable assistance, this work would not have been finished.

I am also indebted to Hijas de Jesús School, especially to Isabel Canal, who cooperated and supported this study.

## **9. References**

Adams, R. (2003). L2 output, reformulation and noticing: Implications for IL development. *Language Teaching Research*, 7(3), 347-376.

- Araceli, C. (n.d.). IRubric: Midterm A1 written rubrics. Retrieved April 26, 2017, from <https://www.rcampus.com/rubricshowc.cfm?code=E24W85&sp=yes&>
- Azkarai, A., & García Mayo, M. P. (2015). Task-modality and L1 use in EFL oral interaction. *Language Teaching Research, 19*(5), 550-571.
- Coyle, Y., & Roca de Larios, J. (2014). Exploring the role played by error correction and models on children's reported noticing and output production in a L2 writing task. *Studies in Second Language Acquisition, 36*(03), 451-485.
- García Mayo, M. P., & Loidi, U. (in press). The Use of Models as Written Corrective Feedback in English as a Foreign Language (EFL) Writing. *Annual Review of Applied Linguistics*, 1-18.
- Martínez Esteban, N., & Roca de Larios, J. (2010). The use of models as a form of written feedback to secondary school pupils of English. *International Journal of English Studies, 10*(2), 143-170.
- Hanaoka, O. (2006). Noticing from models and reformulations: A case study of two Japanese EFL Learners. *Sophia linguistica: working papers in linguistics, 54*, 167-192.
- Hanaoka, O. (2007). Output, noticing, and learning: An investigation into the role of spontaneous attention to form in a four-stage writing task. *Language Teaching Research, 11*(4), 459-479.
- Hanaoka, O., & Izumi, S. (2012). Noticing and uptake: Addressing pre-articulated covert problems in L2 writing. *Journal of Second Language Writing, 21*(4), 332-347.
- Hanjani, A. M., & Li, L. (2014). Exploring L2 writers' collaborative revision interactions and their writing performance. *System, 44*, 101-114.
- Kang, E. Y. (2010). Effects of output and note-taking on noticing and interlanguage development. *TESOL & Applied Linguistics, 10*(2), 19-36.
- Mulligan, C., & Garofalo, R. (2011). A collaborative writing approach: Methodology and student assessment. *The Language Teacher, 35*(3), 5-10.

- Muñoz, C., & Singleton, D. (2011). A critical review of age-related research on L2 ultimate attainment. *Language Teaching*, 44(01), 1-35.
- Neumann, H., & McDonough, K. (2015). Exploring student interaction during collaborative prewriting discussions and its relationship to L2 writing. *Journal of Second Language Writing*, 27, 84-104.
- Robinson, P. (1995). Attention, memory, and the “noticing” hypothesis. *Language learning*, 45(2), 283-331.
- Schmidt, R. (1990). The role of consciousness in second language learning. *Applied Linguistics*, 11, 129-158.
- Schmidt, R. (2010). Attention, awareness, and individual differences in language learning. *Perspectives on individual characteristics and foreign language education*, 6, 27-48.
- Shin, S. Y., Lidster, R., Sabraw, S., & Yeager, R. (2016). The effects of L2 proficiency differences in pairs on idea units in a collaborative text reconstruction task. *Language Teaching Research*, 20(3), 366-386.
- Sotoudehnama, E., & Pilehvari, A. (2016). The impact of peer review on EFL learners' writing proficiency: global and local aspects. *Porta Linguarum: revista internacional de didáctica de las lenguas extranjeras*, (25), 35-47.
- Storch, N. (1999). Are two heads better than one? Pair work and grammatical accuracy. *System*, 27(3), 363-374.
- Storch, N. (2005). Collaborative writing: Product, process, and students' reflections. *Journal of second language writing*, 14(3), 153-173.
- Storch, N. (2013). *Collaborative writing in L2 classrooms* (Vol. 31). Multilingual matters.
- Swain, M. (1985). Communicative competence: Some roles of comprehensible input and comprehensible output in its development. *Input in second language acquisition*, 15, 165-179.

- Swain, M., & Lapkin, S. (1998). Interaction and second language learning: Two adolescent French immersion students working together. *The modern language journal*, 82(3), 320-337.
- Villarreal, I., Gil Sarraeta, N. (submitted). Collaborative writing in secondary school EFL writing: Product and process. *Language Teaching Research*. 1-37
- Wakabayashi, R. (2008). The effect of peer feedback on EFL writing: Focusing on Japanese university students. *On CUE Journal*, 2(2), 92-110.
- Wigglesworth, G., & Storch, N. (2009). Pair versus individual writing: Effects on fluency, complexity and accuracy. *Language Testing*, 26(3), 445-466.
- Williams, J. (1999). Learner-generated attention to form. *Language Learning*, 49, 583-625.
- Yang, L., & Zhang, L. (2010). Exploring the role of reformulations and a model text in EFL students' writing performance. *Language Teaching Research*, 14(4), 464-484.

## 10. Appendix

### 1. Instructions for the first individual informal letter

<b>WRITING TASK: EMAIL</b>
Your school has organized an exchange with an Irish school for next year. Now, you need to be in contact with the students that are going to come to your house in Pamplona. You have to explain them about your place!
Write an email of between 80 and 100 words to the Irish students telling the following:
<ul style="list-style-type: none"> <li>- How does your school look like? Describe it briefly.</li> <li>- Why do you think they are going to enjoy it? What are the things of your school that you like most?</li> <li>- Is there anything that you don't like? Why?</li> </ul>
From:
To: exchange@kylemore.school.ie
Matter: exchange programme

## 2. Model of writing task

Hi,

Let me tell you all about my school, it is called Hijas de Jesús in Spanish. There are girls and boys in the school. The school is semi-private and it is close to the old part of Pamplona. There are three floors. The staff room, the assembly and the nursery school are on the ground floor. Second, fifth and sixth year are on the first floor. All of the primary classes are on the second floor and there are two secondary classes on the top floor (third and fourth). The first years are in a different building. Our four first-year rooms are special. We have a door which opens between us and another class and for half of our timetable we work as one big class. We also have a large sports hall and a big playground.

I think that you'll have a great time here. When we are together in one big class we do a lot of projects together. We make lots of models, create mind-maps, make videos and have a lot of fun. I like history the most so this year I love learning all about Egypt and the Stone Age.

I don't like Maths too much though, the teacher is very nice but I don't understand it very well.

Looking forward to seeing you here

## 3. Instructions for the delayed task

### WRITING TASK (2): EMAIL

Your school has organized an exchange with an Irish school for next year. You are in Ireland now, and you have decided to write to your parents to tell them your experience.

Write an email of between 80 and 100 words telling your family the following:

- ? Describe the Irish school briefly.
- ? Are you having a good time? What things have you learnt?
- ? Is there anything that you don't like? Why?

From:

---

To:

---

Matter:

---

#### 4. Rubric for the assessment of the writings

	Beginning (1) 1 pts	Approaching (2) 2 pts	Meeting (3) 3 pts	Exceeding (4) 5 pts
<b>Content &amp; Development</b>	<p><b>Beginning (1)</b></p> <ul style="list-style-type: none"> <li>- Content is incomplete.</li> <li>- Contains few of the elements required.</li> <li>- Content and purpose of the writing is unclear.</li> <li>- Understanding the intention of the message requires major effort from the reader. Questions were not adequately answered.</li> </ul>	<p><b>Approaching (2)</b></p> <ul style="list-style-type: none"> <li>- Content is not quite comprehensive and lacks accuracy.</li> <li>- Contains only half of the elements required.</li> <li>- Content and purpose of the writing are not very clear.</li> <li>- Understanding the intention of the message requires effort from the reader.</li> </ul>	<p><b>Meeting (3)</b></p> <ul style="list-style-type: none"> <li>- Content is comprehensive and accurate most of the time.</li> <li>- Contains almost all the elements required.</li> <li>- Content and purpose of the writing are clear most of the time.</li> <li>- Understanding the intention of the message requires little effort from the reader.</li> </ul>	<p><b>Exceeding (4)</b></p> <ul style="list-style-type: none"> <li>- Content is fully comprehensive and accurate</li> <li>- Contains all the elements required.</li> <li>- Content and purpose of the writing are clear (a letter to a penfriend).</li> <li>- Understanding the intention of the message requires no effort from the reader.</li> </ul>
<b>Organization &amp; Structure</b>	<p><b>Beginning (1)</b></p> <ul style="list-style-type: none"> <li>- It does not look like a letter, has no opening nor ending as a letter should have.</li> </ul>	<p><b>Approaching (2)</b></p> <ul style="list-style-type: none"> <li>- Even though it looks like a letter, lack either the opening or the ending of it.</li> </ul>	<p><b>Meeting (3)</b></p> <ul style="list-style-type: none"> <li>- Even though it looks like a letter, the opening and the ending are not very clear or are misplaced.</li> </ul>	<p><b>Exceeding (4)</b></p> <ul style="list-style-type: none"> <li>- It is clearly organized as a letter with an opening and ending.</li> </ul>
<b>Grammar, Punctuation &amp; Spelling</b>	<p><b>Beginning (1)</b></p> <ul style="list-style-type: none"> <li>- Grammar mistakes really interfere with meaning.</li> <li>- Too many spelling mistakes that makes it difficult to understand the message.</li> <li>- There is no punctuation which makes the message confusing.</li> </ul>	<p><b>Approaching (2)</b></p> <ul style="list-style-type: none"> <li>- Grammar mistakes interfere with meaning most of the time.</li> <li>- Spelling mistakes obscure the meaning of the message.</li> <li>- Lack of punctuation makes the message quite confusing.</li> </ul>	<p><b>Meeting (3)</b></p> <ul style="list-style-type: none"> <li>- Contains more than 3 grammar mistakes, however, they do not really interfere in understanding the message.</li> <li>- More than three spelling mistakes, however, they do not obscure the meaning of the word.</li> <li>- More than two punctuation mistakes or omissions, however, they do not interfere much in the meaning.</li> </ul>	<p><b>Exceeding (4)</b></p> <ul style="list-style-type: none"> <li>- Contains a maximum of 3 grammar mistakes which do not really interfere in understanding the message.</li> <li>- A maximum of three spelling mistakes which do not obscure the meaning of the word.</li> <li>- A maximum of two punctuation mistakes or omissions which do not interfere with the meaning.</li> </ul>
<b>Vocabulary use</b>	<p><b>Beginning (1)</b></p> <ul style="list-style-type: none"> <li>- Inappropriacies in vocabulary use strains the reader and interfere with message conveying.</li> </ul>	<p><b>Approaching (2)</b></p> <ul style="list-style-type: none"> <li>- Inappropriate use of vocabulary, however, they do not interfere much in conveying the message.</li> </ul>	<p><b>Meeting (3)</b></p> <ul style="list-style-type: none"> <li>- Few errors in the use of the appropriate vocabulary.</li> </ul>	<p><b>Exceeding (4)</b></p> <ul style="list-style-type: none"> <li>- Vocabulary is used appropriately most of the time.</li> </ul>