Research Article

Amparo Lázaro-Ibarrola*

Model texts in collaborative and individual writing among EFL children: noticing, incorporations, and draft quality

Abstract: When written corrective feedback is provided via model texts, language learners notice and incorporate features from the models into their subsequent writings. However, little is known about the accuracy of these incorporations or about the impact of model texts on draft quality. Also, model texts have often been implemented with children working in pairs but, to date, studies including individual and collaborative conditions are extremely scarce. This study examines the impact of model texts among 33 EFL children (aged 11–12) divided into a pair ($N = 22$) and an individual ($N = 11$) group. Our findings do not reveal any significant differences between pairs and individuals. The students in both conditions noticed features, mainly lexical. They incorporated around 50% of these features from the model texts into their final drafts, with an accuracy rate of 60%. Regarding draft quality, the second draft was significantly better than the first one when rated holistically.

Keywords: collaborative writing; EFL children; model texts; noticing; written corrective feedback

1 Introduction

In language lessons, writing tasks are usually carried out individually (Storch 2011) and are accompanied by written corrective feedback (WCF) (Bitchener 2016; Polio 2012). The provision of feedback is coherent with SLA findings showing that WCF appears to have a positive effect on the grammatical accuracy of L2 students' writing (Kang and Han 2015). The common approach of treating writing as a

---
*Corresponding author: Amparo Lázaro-Ibarrola**, I-COMMUNITAS Institute for Advanced Social Research, Universidad Pública de Navarra, Campus de Arrosadia, 31006, Pamplona, Spain, E-mail: amparo.lazaro@unavarra.es

solitary task, however, conflicts with the findings of some recent studies that highlight the language learning benefits of collaborative writing (see review in Storch 2013, 2019). When writing together, learners benefit from their engagement in discussions about language form (Storch 2013; Wigglesworth and Storch 2012) and also seem to be able to write better drafts (Wigglesworth and Storch 2012).

With the value of WCF firmly established (see Polio 2012), research has focused on how effective different feedback techniques are in improving learners' writing performance (Ellis 2009; Lyster et al. 2013). Initially, most studies analyzed feedback techniques that were teacher-generated and focused on isolated errors (e.g., explicit corrections or underlining errors). This led some researchers to propose alternative methods that consider the text as a whole (Yang and Zhang 2010) and that promote a type of noticing that is initiated by the students (Hanaoka and Izumi 2012). Two main techniques comply with this comprehensive and student-centered approach: reformulations, which consist of rewritten versions of students' texts (Cohen 1983), and model texts, which provide students with a model text for comparison (Hanaoka 2007). While both techniques have been shown to effectively promote noticing of linguistic forms, only model texts have an additional advantage: they are easy to implement by teachers (Hanaoka and Izumi 2012). This is one reason why this technique was selected for the present study, given that its feasibility will allow us to recommend (or not) the use of model texts in classroom contexts.

Empirical studies focusing on the use of model texts have been carried out with adults (Hanaoka 2007; Hanaoka and Izumi 2012); adolescents (García Mayo and Loidi Labandibar 2017; Martínez Esteban and Roca de Larios 2010); and children (Cánovas Guirao et al. 2015; Coyle and Roca de Larios 2014; Luquin and García Mayo 2020, 2021). These studies agree that learners (a) notice new lexis and chunks of language from the model texts, and (b) incorporate these noticed features in subsequent drafts. However, there are still some gaps that need to be addressed in order to further our understanding of the potential of model texts as a WCF. For instance, little is known about the accuracy of the features that students incorporate in their drafts. Likewise, there is still no consensus on whether the quality of the writings in which students have incorporated features from the model texts is higher than that of their original writings. In other words, when learners incorporate the noticed features from the model texts, do they incorporate them correctly, or do their texts contain errors (Cánovas Guirao et al. 2015)? Is the final draft in which the features have been incorporated of better quality than the original one (Coyle and Roca de Larios 2014)? In addition, there is little information on the differences between individual and pair work when model texts are used. Previous research has investigated the work of writers working either individually or in pairs; both conditions have rarely been included in the same study (Martínez

Esteban and Roca de Larios 2010). Therefore, studies including both conditions are necessary to identify the specific potential of collaborative writing through model texts.

Within this context, our study partly replicates previous research in that we analyzed the effects of model texts on the noticing and incorporation skills of 33 children learning EFL in primary school (ages 10–11). In order to address the gaps described above, we also analyzed the accuracy of the incorporations and the differences in quality between the initial and the final drafts, measured according to complexity, accuracy, and fluency (CAF) criteria and by a holistic rating scale. Finally, in order to better understand the potential of collaboration, the study included a group of children writing individually ($N = 11$) and a group of children writing in pairs ($N = 22$). Following Storch (2011), collaborative writing is operationalized as the co-construction or co-authorship of a text (Storch 2011). Both participants make decisions together about *what* to write and *how* to write, and when the collaborative writers compare their drafts to the model texts, they notice and discuss the features with one another.

2 Literature review

2.1 Model texts

Model texts are texts written by experts but tailored to students' proficiency and maturity levels. They provide a wide range of context-appropriate L2 words and structures, which help students to become aware of all aspects of language (grammar, lexis, discourse, and so on). Model texts also help students to become aware of new ideas and, at the same time, provide the language to express these new ideas (Coyle and Cánovas Guirao 2019). Thus, model texts demand deeper processing than direct corrections, help learners to generate their own noticing, and promote noticing beyond isolated errors (Hanaoka 2007; Hanaoka and Izumi 2012).

In spite of this purported potential and its subsequent pedagogical appeal, few studies have examined the role of model texts in SLA; similarly, model texts are yet to make their way into language lessons on a large scale (Coyle et al. 2018). To date, model texts have been explored with Japanese (Hanaoka 2007; Hanaoka and Izumi 2012) and Chinese (Yang and Zhang 2010) university students and with Spanish students in primary (Cánovas Guirao et al. 2015; Coyle and Roca de Larios 2014; Luquin and García Mayo 2020, 2021) and secondary schools (García Mayo and Loidi Labandibar 2017; Martínez Esteban and Roca de Larios 2010). The effectiveness of model texts has been tested with students receiving instruction on their use (Coyle

et al. 2018; García Mayo and Loidi Labandibar 2017) and with students in different instructional settings (EFL vs. CLIL) (Coyle and Roca de Larios 2020). In relation to other WCF strategies, the use of model texts has been compared to reformulations with adults (Hanaoka and Izumi 2012; Yang and Zhang 2010); same-task repetition (Cánovas Guirao); self-editing (Luquin and García Mayo, 2021); and error correction (Coyle and Roca de Larios 2014) with children. Regarding collaboration, model texts have been implemented with students working in pairs (Cánovas Guirao; Coyle and Roca de Larios 2014; Luquin and García Mayo 2020, 2021) and with students working individually (García Mayo and Loidi Labandibar 2017; Hanaoka 2007; Hanaoka and Izumi 2012). Rarely, however, have individual work and pair work been investigated in the same study; to the best of our knowledge, only one small-scale study has included both conditions (Martínez Esteban and Roca de Larios 2010).

2.1.1 Model texts, noticing, and incorporations

The main focus of the studies on model texts has been on their potential to promote noticing (Schmidt 1990). This noticing has been operationalized as (a) the features that students report while writing an initial draft (referred to in the literature as problematic features noticed, or PFNs), and (b) the features they report while comparing their draft to the model texts (referred to as features noticed, or FNs). The PFNs and FNs have mostly been identified by asking students to write them down (note taking); in some studies which included pairs, they were also coded using the recordings of students' oral interactions (Cánovas Guirao et al. 2015; Luquin and García Mayo 2020, 2021; Yang and Zhang 2010).

With the FNs and PNs well identified, researchers then examined whether they were incorporated in the students' final drafts, taking into account the quantity and type of incorporations and, in some cases, also explaining whether these incorporations were linked to the PFNs, to the FNs, or to both (Coyle and Roca de Larios 2014; Hanaoka 2007; Hanaoka and Izumi 2012). These studies have shown that, when learners compare their writings to model texts, they predominantly notice lexical features and are able to incorporate these in their subsequent writings. Hanaoka (2007), in particular, highlights the specific potential of a good model to help learners identify the linguistic features they need, which, in turn, increases the learners' opportunities to use and learn these features.

With regard to how the incorporations have been treated in the literature, most researchers include them in their accounts, even when they contain "minor errors" (Coyle and Roca de Larios 2014; García Mayo and Loidi Labandibar 2017; Hanaoka 2007; Martínez Esteban and Roca de Larios 2010). Only the study by Cánovas Guirao et al. (2015) considers the acceptability of students' incorporations.

Cánovas Guirao et al. (2015) analyzed 10 pairs of children divided into two groups, a model group and a task repetition group (control group). In addition to examining noticing and incorporations, these authors classified the modifications as acceptable and unacceptable (containing lexical or grammatical errors) and found a majority of acceptable changes in both groups, although the number was greater in the model group (64%). These acceptable incorporations were mainly lexical and most were directly or indirectly related to the model texts.

Nevertheless, while research has concentrated on reporting the noticing of features and subsequent incorporations, little is known about the overall impact that working through model texts has on the quality of the students' writing. With this in mind, Coyle and Roca de Larios (2014) compared the grammaticality of the initial and final draft of 46 Spanish learners of English (ages 11–12) working in pairs in two groups. One group received feedback via model texts and the other via direct error correction provided by the teacher. Coyle and Roca de Larios used a coding scheme, which identified three levels of acceptability and comprehensibility: (a) preclauses, which are incorrect and with unclear meaning; (b) protoclauses, which are incorrect but with clear meaning; and (c) clauses, which are mainly correct and with clear meaning. Their findings showed that the second draft was indeed better: it had a greater number of clauses (and fewer preclauses). The improvement was identified in both conditions but was greater in the error correction group than in the model text group. However, it is important to note that these authors accepted "slight inaccuracies in spelling, lexis, grammar or concordance within the category of 'clauses'" (Coyle and Roca de Larios 2014: 463) and, therefore, more thorough analyses of accuracy are still needed.

In sum, the scarce yet positive results when accuracy has been considered (Cánovas Guirao et al. 2015; Coyle and Roca de Larios 2014) pave the way for future research on model texts: it is necessary to further examine the accuracy of the incorporations and, in turn, the quality of the final draft.

2.1.2 Model texts and collaboration

Collaborative writing is defined as "an activity that requires the co-authors to be involved in all stages of the writing process, sharing the responsibility for and the ownership of the entire text produced" (Storch 2019: 40). Several studies have demonstrated that collaborative writing can be considered an effective practice (García Mayo 2021; Storch 2011, 2019; Wigglesworth and Storch 2012). Firstly, it offers learners the benefits attributed to writing *per se*, considering writing within the output hypothesis (Swain 1985, 2000) and the writing-to-learn approach (Manchón 2011). Secondly, it combines oral and written modalities, offering students the language learning benefits associated with both writing (e.g., slow pace,

visible output) and speaking tasks (Storch 2021). Finally, with regard to motivation, learners express positive feelings about collaborative writing activities (Shehadeh 2011; Storch 2005, 2013), while they describe individual writing as boring (Murtiningsih 2016). As recently suggested by Neumann and McDonough (2014, 2015) in their studies on collaborative prewriting tasks, collaboration may be beneficial in terms of text quality, even for students who prefer individual work (Neumann and McDonough 2014, 2015).

These benefits are supported by a large number of empirical studies showing that writing pairs generate better texts than individual writers (McDonough and García Fuentes 2015) and that learners benefit from the language discussions they engage in while composing their drafts (Fernández Dobao 2012; Storch and Wigglesworth 2009; Villarreal and Gil-Sarratea 2020). For instance, Storch and Wigglesworth (2009) compared the writings of pairs ($N = 48$) and individuals ($N = 48$) in an Australian university and identified improvements in accuracy in the compositions written by the pairs. When analyzing pair talk, they found that over 30% of the discussions focused on language issues, which these authors interpret as suggestive of the usefulness of collaborative writing activities. In line with these findings, Fernández Dobao (2012) analyzed the writings of university learners of Spanish in the US, writing individually ($N = 21$), in pairs ($N = 15$), and in groups of four ($N = 15$). She found that the groups wrote the most accurate texts, followed by the pairs. She also analyzed the language-related episodes (LREs) generated by the pairs and the groups. LREs are defined as "any part of a dialogue where the students talk about the language they are producing, question their language use, or correct themselves or others" (Swain and Lapkin 1998, p. 326). Her results revealed that the groups produced more LREs and a higher percentage of correctly resolved ones. Also confirming the superiority of collaboration, Villarreal and Gil-Sarratera (2020) conducted a study with Spanish EFL learners in secondary school, divided into a control group ($N = 16$) writing individually and an experimental group ($N = 16$) writing in pairs. They found that the pairs produced better texts and were able to pool ideas, discuss linguistic issues, and provide feedback to each other.

In the case of model texts, data has been collected from students writing individually (García Mayo and Loidi Labandibar 2017; Hanaoka 2007; Hanaoka and Izumi 2012); in pairs (Cánovas Guirao et al. 2015; Coyle and Roca de Larios 2014, 2020; Luquin and García Mayo 2020, 2021); or in pairs but writing the final composition individually (Yang and Zhang 2010). This leaves a clear gap of research including both conditions (pair vs. individual) within the same study (Martínez Esteban and Roca de Larios 2010).

The studies that have included both collaboration and model texts have mainly been carried out with YLs. As explained when dealing with model texts and noticing, Coyle and Roca de Larios (2014) compared the use of explicit corrections

and model texts in 23 proficiency-matched pairs of primary school children (ages 11–12), and Cánovas Guirao et al. (2015) compared same-task repetition and model texts with two groups of 10–11-year-olds ($N = 20$). Their results regarding the model condition showed that the pairs were able to successfully notice and incorporate lexis and chunks of language. More recently, Luquin and García Mayo (2020, 2021) also implemented model texts in collaboration and compared pairs of 11–12-year-olds working through model texts versus pairs self-editing their texts. They found that the model text group was able to notice more lexical and content-related features and to incorporate more mechanics- and discourse-related features than the self-editing group. Finally, Coyle and Roca de Larios (2020) explored the effect of the instructional setting (EFL vs. CLIL) in pairs of primary school students (ages 9–11) working through model texts; they found that all pairs noticed surface differences, and the CLIL pairs also paid attention to new and alternative content.

However, to the best of our knowledge, there is only one study which includes both pairs and individuals: Martínez Esteban and Roca de Larios (2010). This small-scale study analyzed data from 17 to 15-year-olds learning EFL in secondary school. Five participants worked individually and 12 worked in pairs through a three-stage writing task: they wrote compositions based on a picture prompt, compared them with model texts, and finally, in a subsequent session, rewrote their initial drafts. Results showed that the students noticed mainly lexical problems, but they only found a few solutions for them in the model texts. Also, the students noticed a large number of features related to the content and means to express it, and they were able to incorporate a large number of these in their revised texts. This was especially true in the case of the pairs. These positive but insufficient results encourage more research on model texts comparing pairs and individuals.

2.1.3 Summary of literature review and impetus for the present study

In sum, the need for the present study on the effectiveness of model texts among EFL children is sustained in several research gaps that have been described above and are summarized here.

To begin with, the potential of model texts to promote noticing and incorporations has been firmly established. However, little is known about the accuracy of the incorporations (Cánovas Guirao et al. 2015) or about the differences between the students' initial and final drafts (Coyle and Roca de Larios 2014). In our study, and in line with previous research, we analyzed the noticing of mismatches between students' drafts and model texts, and the subsequent incorporation of this noticing. In addition, and in order to take a step forward into understanding model texts, we also looked into the accuracy of the incorporations and into the quality of the initial vs. final drafts in order to see whether students'

final drafts, with incorporations from the model texts, were of higher quality than their initial drafts.

Second, little is known about whether there are differences in the effectiveness of model texts when we compare students working individually with students in pairs. Our study included both conditions and thus adds to our understanding of this issue. As mentioned above, to the best of our knowledge, only one small-scale study has included both conditions (Martínez Esteban and Roca de Larios 2010). Our findings could thus help to further understanding of collaboration in general. While studies seem to unanimously show that collaboration is positive in oral tasks, the conjoint potential of writing and collaboration still remains a relatively under-researched arena (Storch 2016) and a practically unexplored one if we think of young second language learners (Coyle and Roca de Larios 2014).

At the same time, our study also bears pedagogical interest. We tested two activities that are suitable for classroom contexts: collaborative writing and the provision of WCF via model texts. This means that our findings could encourage the implementation of these two activities in language lessons. Finally, from a general SLA perspective, our study adds to our understanding of YLs, a large and growing population that is still relatively under-researched (Collins and Muñoz 2016).

3 Research questions

The following research questions were addressed in our study:
1. What aspects of language do pairs and individuals notice when writing (PFNs) and comparing their own text to a model text (FNs)?
2. What do pairs and individuals incorporate into their final writings? Are these incorporations linked to previously noticed features, that is, to the PFNs while writing and/or to the FNs when comparing their writing to the model texts?
3. Are the incorporations introduced by the pairs and individuals accurate?
4. Is the draft produced by the pairs and individuals after the comparison session better than the learners' original draft?

4 Method

4.1 Participants

The participants in this study were 33 children, 18 boys and 15 girls, aged 10–11, from two EFL classes in the same primary school in northern Spain. The children had been learning English in the school since the age of three and had received 4 h of English lessons per week every school year. The method followed was

communicative, and the teaching was mainly guided by textbooks from well-known publishers. Before starting with the project, we obtained written permission for the learners' participation from the school and from the children's parents or guardians. As the tasks were carried out during students' regular lessons, and to respect the ethics of the school, the students who did not grant permission also participated in the activities but their production was not considered for research purposes.

The 33 children were divided into two groups, one in which students worked individually ($N = 11$; 5 boys and 6 girls; mean age 11) and 1 in which students worked in pairs ($N = 22$; 13 boys and 9 girls; mean age 10.95). These groups will be referred to as the individual group (henceforth IG) and the collaborative group (CollG).

Prior to data collection, the proficiency of the IG and the CollG was measured by an English test (Cambridge A2 Flyers). The test was chosen with the help of the students' teacher because it matched the level of the group and because the students were familiar with its format. Based on the results obtained, a U-Mann–Whitney analysis showed that the two groups displayed a normal distribution in their scores ($p = 0.887$). Thus, an independent-samples t-test was conducted to compare the proficiency scores in the IG and CollG. There was no significant difference between the two groups (IG M = 6.82; SD = 1.9; CollG M = 6.91; SD = 1.3; $t(29) = 0.152$; $p = 0.880$).

4.2 Pilot task

The children were used to pair and teamwork in their English lessons; however, they had never written a text collaboratively before, they had never written a narrative based on a picture prompt, and they had never received WCF in the form of model texts. In order to familiarize them with the task and procedure and following previous studies (Coyle and Roca de Larios 2014), a pilot task was administered to the students three weeks prior to data collection. To this end, we conducted two training sessions with the teacher to make sure she understood the task and the procedure to follow. Later, the teacher taught the students in one of her regular school sessions. In this session, all students wrote a composition in pairs based on a picture prompt, compared it to two model texts, and were asked to notice any difficulties they encountered while writing and comparing. One researcher was also present in the classroom when the pilot task was administered. The teacher explained to the students that writing collaboratively meant that they were a team: they had to discuss what they wanted to write and how they wanted to write it. Likewise, the teacher told the students that they had to discuss the differences between the model text and their draft.

4.3 Data collection

Following common practice in studies on model texts since Hanaoka's (2007) study, a set of pictures was used to provide students with a writing prompt. When using picture prompts the content of the writing is constrained by the visual stimuli, and this ensures the comparability of the students' drafts and the model texts. The authors worked with the classroom teacher to choose two picture prompts to form the basis of students' writing. Two sets of pictures presenting a clear narrative story line with familiar content were selected to make sure the students could accomplish the task. The picture prompts had five vignettes and belonged to the Cambridge Young Learners English Test. One picture prompt was used during the training sessions to familiarize students with the task, and the other was used to collect the data for the present study (see Appendix 1). The teacher agreed that, in terms of content, the plot of the stories depicted in the vignettes would be understood by the students. In terms of language, the teacher explained that the pictures included some known vocabulary and some challenging words. All of these factors made them suitable to use as writing prompts.

The data collection process was carried out over three sessions (one per week) in November 2017:

Stage 1. Composing. Each pair in the CollG and each individual in the IG were provided with a sheet (Sheet 1) on which they had a set of pictures and instructions to write their composition and to note any problems they experienced or concerns they had while writing (see Appendix 1). Thus, noticing was measured by means of note taking in both groups (Hanaoka 2007). Taking notes is quite demanding, and students might notice more than they report; nonetheless, learners' notes seem to be a very good indicator of the focus of learners' attention (Hanaoka 2007). The draft produced in this session will be referred to as Draft 1, and the notes, which are used as evidence of students' noticing, will be referred to as PFNs. At the end of this session, Sheet 1, containing Draft 1 and the PFNs, was collected.

Stage 2. Comparison. A week after composing, the pairs and the students working individually received Sheet 1 with their original compositions (Draft 1) and their PFNs, and were also given Sheet 2 (see Appendix 2). Sheet 2 contained two model texts and instructions asking students to compare these texts to their own compositions, write down the differences, and explain whether they had learned something from these model texts. These notes will be referred to as FNs. The model texts had been written by two native speakers, one from the Republic of Ireland and one from the UK, who were expert EFL teachers in secondary and primary education in Spain and who were, therefore, well acquainted with the genre, proficiency, and cognitive level of the participants. In line with previous research, a week passed between Stages 1 and 2, and two model texts were used to

avoid memorization or copying from a single text (Coyle and Roca de Larios 2020; Hanaoka 2007). At the end of this session, Sheet 1 and Sheet 2 were collected.

Stage 3. Rewriting. One week later, both groups were given Sheet 3. This sheet simply contained the same picture prompt used in Stage 1 and instructions to write another composition based on the pictures, this time trying to incorporate the features they had noticed from the model texts. We will refer to this second composition as Draft 2. Following previous research on model texts (Coyle et al. 2018; Coyle and Roca de Larios 2020; Hanaoka and Izumi 2012; Luquin and García Mayo 2020, 2021; Cánovas Guirao et al. 2015), we used the same picture prompt, and the students did not have access to their original draft (Draft 1) when writing Draft 2. By using the same picture prompt, we wanted to facilitate the inclusion of the new content and new forms of expression they had noticed. By not giving students their original draft and notes, we wanted to avoid mere editing or copying.

With regard to timing, in line with previous studies comparing collaborative and individual writing (Wigglesworth and Storch 2009), the CollG was given more time than the IG. Thus, the CollG was given 30 min to complete each of the tasks and the IG 20 min. None of the participants had difficulties in completing the task within the allotted time.

The pairs performed the task in the computer room of the school. The room was big enough to allow five pairs to carry out the task at the same time. A researcher was with them at all times, but she did not provide any help with the task. The individual writers were in their own classroom with the other researcher and, as in the case of the pairs, she simply asked the students to perform the task and did not provide any help.

4.4 Data analysis

The analysis carried out to answer the research questions includes the following set of data from each participant in the IG and from each pair in the CollG: Draft 1 and PFNs from Stage 1, FNs from Stage 2 and, finally, Draft 3 from Stage 3. In total, we collected 22 original stories (Draft 1) and notes (PFNs); 22 sets of notes from comparison (FNs); and 22 final stories (Draft 2).

Our analysis includes the features noticed (PFNs and FNs), the incorporation of these features in Draft 2, and a comparison of the quality of the initial (Draft 1) and final (Draft 2) compositions.

4.4.1 Noticed features and incorporations

Following Coyle and Roca de Larios (2014) and Hanaoka (2007), all noticed features were classified into the following five categories: lexis, spelling, grammar,

content, and other. Table 1 briefly defines the scope of these categories and illustrates them with examples from our database. As all extracts were transcribed literally, the examples include some spelling errors in Spanish.

Table 1: Categories of noticed features.

Category	Definition and examples
Lexis	**Students report issues related to vocabulary and lexical choices:** ***Example 1.*** No se decir enfadado y no he puesto nada (PFN, I3). (*I don't know how to say angry and I have not written anything*).
Spelling.	**Students report issues related to the spelling of words:** ***Example 1.*** No nos acordábamos como se escribia "scissors" y ahora sí (FN, P4). (*We didn't remember how to write "scissors" and now we do*)
Grammar	**Students report issues related to syntax and morphology (tenses, agreement, etc.):** ***Example 1.*** Nos hemos dado cuenta de que era mejor poner "his" (FN, P6). (W*e noticed that it was better to write "his"*)
Content	**Students report noticing of ideas or expressions from the model texts.** ***Example 1.*** Me ha gustado la frase "proud of her work" (FN, I7). (*I liked the sentence "proud of her work"*). ***Example 2.*** No hemos puesto que estaban en el "sitting room" (PFN, P1). (*We did not write that they were in the "sitting room"*).
Other noticing	**This category comprises the students' notes that did not fall in any of the other categories. In our data base, the instances that did not fall in the previous categories referred to students' expressions of their personal opinions on the model texts and on their own writings:** ***Example 1.*** Nos a gustado mas el modelo A porque esta mas completo (FN, P11). (*We prefer model A because it is more complete*). ***Example 2.*** La redacción de los chicos ingleses está mejor que la mía (PFN, I1). (*The composition by the English kids is better than mine*). ***Example 3.*** Me ha gustado mucho el modelo A y el B porque han dicho lo mismo pero cada uno de una forma diferente (PFN, I3). (*I liked models A and B a lot because they say the same but in a different way*). ***Example 4.*** Me gusta más el modelo A porque tiene más vocabulario (FN, I2). (*I prefer model A because it has more vocabulary*).

The PFNs while writing Draft 1 were also classified as solvable if a plausible solution could be found in the model texts, and unsolvable if there was no solution in the model texts, as Table 2 illustrates.

Table 2: Examples of solvable and non-solvable features.

PFNs	Extract from model	Classification
P3: We don't know how to write "tijeras" (scissors)	"…she gets some **scissors** and glue, cuts them out and sticks them onto her picture"	Solvable
P1: "Estantería" no sabíamos y no lo hemos puesto (We didn't know how to say "shelf" and we haven't written it)	The word "**shelf**" does not appear in the models	Not solvable

The noticed features were tracked to the students' final drafts and classified as incorporated if students included them in their final draft, and unincorporated if they did not. Then, the incorporated features were further classified (see Table 3) into correctly or incorrectly incorporated. Only fully correct incorporations were considered as correctly incorporated; that is, all minor errors, including errors in spelling, were classified as incorrectly incorporated. For example, when learners incorporated the word "scissors" with correct spelling and meaning, this feature was classified as correctly incorporated; in contrast, when the students incorporated a feature from the model texts but with errors, for example when the word "scissors" was incorporated as "siccors," the feature was classified as incorrectly incorporated.

Table 3: Examples of correctly and incorrectly incorporated features.

	Incorporation	Accuracy	Extract from Draft 2
FNs & PFNs	Incorporated	Correctly	She take the glue and the <u>scissors</u> (I1)
		Incorrectly	She go to take the <u>siccors</u> (P3)
	Not incorporated	The noticed feature does not appear in Draft 2	

We need to clarify that within the category of incorporated features we also included features related to content because, although introducing content *per se* is not related to accuracy, these incorporations could be correct or incorrect. That is, when a pair reported information such as "The model says *It's Monday*" they could incorporate this opening into their composition correctly or incorrectly (e.g.,

"is Monday"). However, the features related to the "other noticing" category were not considered as they related to students' opinions on the model texts and, therefore, could not be incorporated (e.g., "I liked models A and B a lot because they say the same but in a different way"; "The composition by the English kids is better than mine" (see Table 1).

We also examined whether incorporated features had been noticed in Stage 1, in Stage 2, or in both, that is, if these incorporations corresponded to PFNs, FNs, or both. Table 4 features one example of each (top to bottom).

Table 4: Incorporations and noticing stages.

	Incorporation from Stage 1 (Example from P8)	Incorporation from Stage 2 (Example from P3)	Incorporation from Stages 1&2 (Example from P9)
Draft 1	She cut the flowers and put in the picture		the girl take the seasars
PFN	Quería decir "cortó flores de la tarea de su hermano" pero no se si se entenderá (I meant 'she cut the flowers from her brother's homework' but I don't know if it is clear).		He escrito 'seasars' (tijeras) pero no se si se escribe así (I have written 'seasars' but I am not sure that this is how you write it).
FN		No hemos escrito que el chico se había enfadado al final. (We didn't write that the boy got angry at the end).	Y he escrito la palabra tijera así: 'seasors' cuando he visto en los modelos que se escribe 'scissors'. (And I have written the word 'tijera' this way: 'seasors', but I have seen in the models that it is 'scissors').
Incorporation in Draft 2	the girl cut's her flowers of her brother's homework	and he is very angry	and then Katy take the scisors and the glue

4.4.2 Comparison of students' initial and final drafts

The quality of the students' drafts was analyzed using quantitative and qualitative measures. The quantitative measures included complexity, accuracy, and fluency (CAF), measurements that have been used in prior research on individual vs. collaborative writing (Storch 2005; Villarreal and Gil-Sarratea 2020; Wigglesworth

and Storch 2009). Accordingly, the specific elements of analysis for each component were chosen following these studies.

–   Accuracy was measured by calculating the proportion of errors to total number of words and the proportion of error-free clauses to total number of clauses. Lexical errors (word choice) were only considered when the word used obscured meaning. As in previous research, punctuation errors were excluded, but spelling, which constitutes a category of noticing, was considered. Thus, errors belonging to the three following categories were coded: (a) grammatical, syntactical, and morphological errors (e.g., errors in word order, missing elements, verb tense, subject–verb agreement, etc.); (b) lexical errors, which included word choice and L1 transfers; and (c) spelling errors (Fernández Dobao 2012).
–   Complexity was measured as number of clauses per T-unit. In addition to this, we added D as a measurement for specific lexical density (Malvern et al. 2004; McKee et al. 2000).
–   Fluency was measured by considering the total number of words and the proportion of T-units per text and average number of clauses per text.

The qualitative measures included a five-scale global evaluation scheme based on Storch (2005) and adapted to the content of the task (see Appendix 3). This scale, already used in some recent studies using the same picture prompt (Lázaro-Ibarrola and Hidalgo 2021; Hidalgo and Lázaro-Ibarrola 2020), is designed to measure the content, structure, and degree of task fulfillment (Swales and Feak 1994); that is, its purpose is to gauge an overall impression that will help complete the quantitative results.

4.5 Inter-rater reliability

The participants' written production was coded by the author of this paper. An independent research assistant also analyzed the production of six pairs and six individual writers (18% of the data). The research assistant had an academic background in linguistics an
    d EFL teaching and had been working in the field of EFL writing research for one year at the time of data analysis. Both raters held several meetings prior to data coding to agree on their understanding of the measures of analysis. During the meetings they evaluated several samples together. These samples were compositions written by children of a similar age and level and formed part of the database of a larger research project on collaborative writing within which our study was conducted.

After their independent coding of the data presented in this study, the raters met in order to resolve the few discrepancies on a case-by-case basis until they reached an agreement on each one. Total agreement was found for the analysis of features, fluency, and complexity. In the case of accuracy and holistic measures, some discrepancies were found and inter-rater agreement was calculated using Cohen's Kappa. Substantial agreement was found for holistic measures in Draft 1 (k = 0.751) and almost perfect agreement in Draft 2 (k = 0.814, p = 0). As for accuracy, moderate agreement was found in the case of errors per number of words (k = 0.364, p = 0) and substantial agreement in the case of error-free clauses (k = 0.634, p = 0). Discrepancies in accuracy were mainly related to errors classified as word choice.

With regard to statistical analysis, Dependent Samples t-tests were used for data that presented a normal distribution, and Wilcoxon signed-rank tests were used for data that were not normally distributed. The significance level was set at α = 0.05.

5 Results

5.1 Results for noticing and incorporations

The results regarding PFNs in both groups and classified into types are presented in Table 5. The bottom row shows the number of solvable features.

Table 5: Frequencies and proportions of problems noticed in the Stage 1 writing task.

PFNs	Individual				Pairs				All participants			
	N	**%**	**Mean**	**SD**	**N**	**%**	**Mean**	**SD**	**N**	**%**	**Mean**	**SD**
Lexis	12	63.15	1.09	1.3	11	79	1	0.9	23	69.69	1.04	1.1
Spelling	3	15.78	0.27	0.6	0	0	0	0	3	9.09	0.13	0.5
Grammar	1	5.26	0.09	0.3	0	0	0	0	1	3	0.045	0.2
Content	2	10.52	0.18	0.6	2	14	0.18	0.4	4	12.12	0.36	0.5
Other	1	5.26	0.09	0.3	1	7.1	0.09	0.3	2	6.06	0.09	0.3
Total	19	100			14	100	1.27		33	100	1.4	
Solvable	15	78.94			10	71.42			25	75		

As Table 5 shows, the number of PFNs was very low, only 19 in the IG and 14 in the CollG. To the data in this table we need to add that four individual writers and one pair did not provide any features at this stage (see Appendix 4 for summary of

individual results regarding noticing and incorporations). A statistical analysis shows no difference between the two groups (U = 52, p = 0.585). Also, all participants mainly reported lexical features (23/33 = 69.69%), with very few instances of noticing in other categories. Finally, most of these features were solvable: 15 out 19 in the IG and 10 out of 14 in the pairs, amounting to 75%.

Compared to the PFNs (33 in total), the number of FNs through the comparison stage was clearly higher, reaching a total of 100 features (see Table 6). All participants provided at least one instance of noticing at this stage (see Appendix 4 for individual results).

Table 6: Frequencies and proportions of features noticed in the Stage 2 comparison task.

FNs	Individual				Pairs				All participants		
	N	%	Mean	SD	*N*	%	Mean	SD	*N*	%	Mean
Lexis	22	37	2	1.3	21	53	1,909	2.7	43	43	1.95
Spelling	4	6.7	0.364	0.7	0	0	0	0	4	4	0.18
Grammar	2	3.3	0.182	0.4	2	5	0.182	0.6	4	4	0.18
Content	9	15	0.818	1.3	1	2.5	0.091	0.3	10	10	0.45
Other	23	38	2,091	1.9	16	40	1,455	1.5	39	39	1.77
Total	60	100	5,455		40	100	3,636		100	100	4.5

When comparing the two groups, the number of FNs was greater in the individuals than in the pairs although this difference did not reach statistical significance ($t(20)$ = 1.93, p = 0.067); however, given that the number of observed features was so low, we recognize that the power of the statistical measures could be limited (Larson-Hall 2010). When looking at the types of noticing, most FNs were focused on lexical aspects, as was the case for PFNs. Nevertheless, the "other noticing" category, which comprised students' opinions on the model texts became quite important in both groups (e.g., "I prefer model A because it has more vocabulary"; "The composition by the English kids is better than mine" (see Table 1). Also, FNs in relation to content increased among individuals (nine instances).

Next, Table 7 shows the results regarding incorporations. The percentage is calculated over the total number of noticed forms that can be incorporated and, therefore, the noticing related to the "other noticing" category is not considered.

Table 7: Incorporations across Stages and percentage of incorporations over total number of noticed features in Stages 1 & 2.

Incorporations	Individual	Pairs	All participants
Total noticing (FNs & PFNs, excluding the category "other")	37	24	61
Incorporations from Stage 1	0	2	2
Incorporation from Stage 2	12	3	15
Incorporations from Stages 1&2	7	6	13
Total number of incorporations	**19**	**11**	**30**
Percentage of incorporations over total number of noticed features (excluding the category "other")	**51.35**	**45.83**	**48.59**
Correctly incorporated features	**9**	**8**	**17**
Percentage of correctly incorporated features over total number of incorporated features	**47.37**	**72.72**	**60.04**

As Table 7 shows, the percentage of incorporations was around 50% for both groups (individual 51.35, collaborative 45.38) with no statistical differences between them ($t(20) = 1.62$, $p = 0.121$). Three pairs and two individuals did not incorporate any features (see Appendix 4). When looking at whether the incorporations had been previously reported, most came from the FNs at Stage 2 or from both stages (PFNs at S1, which were also FNs at S2). Only two incorporations by the pairs (none by the individuals) were related to features noticed exclusively at Stage 1; both corresponded to solvable features.

Finally, the accuracy of the features incorporated was greater in the pairs, who incorporated eight features correctly and only three inaccurately, while the individuals incorporated nine features correctly and 10 incorrectly. However, the differences between the correctly and incorrectly incorporated features in the two groups did not reach statistical significance (correct: $U = 55$, $p = 0.693$; incorrect: $U = 38$, $p = 0.085$).

5.2 Results for CAF and holistic ratings

To examine the impact of the model texts on the final draft, the quality of the second draft in relation to the first one was measured according to the complexity, accuracy, and fluency (CAF) framework and by a holistic rating. The results are presented in Table 8.

Table 8: CAF measures.

			Individual	Pairs
Accuracy	**Errors per total number of words**	**Draft 1**	0.31	0.23
		Draft 2	0.26	0.21
		Dependent samples *t*-tests	$t(10) = 1.829$, $p = 0.097$	$t(10) = -0.16$, $p = 0.876$
	Proportion of error-free clauses over total number of clauses	**Draft 1**	0.13	0.21
		Draft 2	0.18	0.24
		Wilcoxon signed-rank tests	$Z = -1.41$, $p = 0.158$	$Z = -1.51$, $p = 0.131$
Complexity	**Lexical density**	**Draft 1**	48.53	47.92
		Draft 2	52.81	48.97
		Dependent samples *t*-tests	$t(10) = -1.909$, $p = 0.811$	$t(10) = -0.246$, $p = 0.85$
	Clauses per T-unit	**Draft 1**	1.11	1.12
		Draft 2	1.10	1.13
		Dependent samples *t*-tests	$t(10) = 0.338$, $p = 0.742$	$Z = -0.8, p = 0.424$
Fluency	**Total number of words**	**Draft 1**	795	888
		Draft 2	739	939
		Dependent samples *t*-tests	$t(10) = 1.316$, $p = 0.218$	$t(10) = -1.777$, $p = 0.218$
	Proportion of T-units per text	**Draft 1**	105 (9.5)	109(9.9)
		Draft 2	103 (9.3)	117 (10.6)
		Dependent samples *t*-tests	$t(10) = 0.377$, $p = 0.714$	$Z = -1.136$, $p = 0.256$
	Number of clauses per text	**Draft 1**	117 (10.6)	122 (11.09)
		Draft 2	113 (10.27)	132 (12)
		Dependent samples *t*-tests	$t(10) = 1.174$, $p = 0.267$	$Z = -1.488$, $p = 0.137$
Holistic rating (1–5 Scale)		**Draft 1**	2.36	2.90
		Draft 2	3	3.72
		Dependent samples *t*-tests	$t(10) = -3.464$, $p = 0.006$	$Z = -2.530$, $p = 0.011$

Starting with the results obtained for CAF, the percentages showed heterogeneous results for each measurement. Some findings suggesting a tendency to greater accuracy and lexical density in Draft 2 (particularly in the individuals) approached but did not reach significance. Structural complexity (clauses per T-unit) and also fluency seemed to be quite stable across drafts, showing only small fluctuations. Nevertheless, none of the analyses carried out for CAF measures showed statistically significant differences when Draft 1 and Draft 2 were compared.

The holistic analysis, however, clearly showed that both groups significantly improved from Draft 1 to Draft 2; that is, the second draft, when evaluated holistically, seemed to be better than the first draft in both conditions. This is particularly so in the case of the pairs, where the degree of improvement ranged from 2.90 to 3.72.

6 Discussion

The aim of the present study was to investigate the potential of model texts as a WCF technique to promote the noticing and incorporation of linguistic features and, as a result, to assess the potential of model texts to help students generate a text of better quality. The study was carried out with YLs of English in a school context; one group worked individually and one group worked in pairs in order to explore the specific potential of model texts in the context of collaborative writing among children.

6.1 Writing stage: problematic features noticed

The results regarding noticing show that the number of PFNs while writing was very low (33 in total); these were mainly lexical (23 out of 33), and no significant differences were seen between the IG and CollG. When comparing the quantity of PFNs with other studies with model texts, where noticing is also codified from note taking, our participants obtained the lowest mean, with only 1.4 instances per student or pair. Going from highest to lowest, Hanaoka and Izumi (2012) reported a rate of 4.7; Coyle and Roca de Larios (2014), who also studied children, reported a rate of 4.25 (calculated from Coyle and Roca de Larios 2014: 464); Hanaoka (2007) reported a mean of 3.5; García Mayo and Loidi Labandibar (2017) reported a mean of 3.11; and, finally, Martínez Esteban and Roca de Larios (2010) reported the closest mean to ours, 2.63.

The lower rate of PFNs found in our study could be attributed to several causes: the younger age and lower proficiency level of our children (Coyle and Roca de Larios 2014); their limited ability to report information in the written form (García Mayo and Loidi Labandibar 2017); and/or the fact that they had only worked in pairs and through model texts one time before the task was administered. Coyle and Roca de Larios (2014), with a study very similar to ours, obtained a much higher mean (4.25). In their study the teacher seems to have trained the participants for the joint story-writing task quite intensely, while in our study only two sessions were devoted to this training. It is possible that, with training, the amount of noticing could be increased (Coyle et al. 2018).

The fact that the PFNs were almost exclusively lexical fully coincides with previous research, which seems to confirm that noticing of lexis and means of expression is the greatest asset of model texts (Coyle et al. 2018; Hanaoka 2007; Luquin and García Mayo 2021). This finding is also in line with the idea that, due to the processing limitations of the human brain, meaning is favored over form; this is particularly the case at low levels of proficiency (Van Patten 1990; Wong 2001).

6.2 Comparison stage: features noticed

At the comparison stage, that is, when students compared their drafts to the model texts, the total number of FNs that students reported was 100. Thus, the number of noticed features while comparing was larger than the number of noticed features while writing (100 FNs vs. 33 PFNs). Most of the FNs reported at the comparison stage were lexical (43, representing 43% of the total number) but, unlike the case of PFNs, a large number (39, representing 39%) fell into the "other noticing" category. In this category, students expressed their preferences and opinions on the model texts using expressions such as "I prefer model A because it contains more information" or "I liked the model texts; they are better than our drafts." These comments also seemed to reveal the students' capacity to see the model text as a whole beyond the chunks of lexis or grammar that they also noticed.

In the case of the individuals, there were also a few FNs focused on content (which represented 15% of the total features in this group). This category corresponded to ideas that were present in the pictures but which the students had not included in their drafts, such as "I didn't say the time" or "I didn't know it was Friday afternoon." This confirms the potential of model texts to help students notice new content while, at the same time, providing the language to express this content (Coyle and Cánovas Guirao 2019).

When looking at the two experimental conditions, the number of FNs at this stage was greater in the individuals than in the pairs (60 vs. 40); however, a statistical analysis showed that this difference was non-significant. Contrary to our expectations and to the findings from the small-scale study by Martínez Esteban and Roca de Larios (2010), in which pairs noticed more, in our study, there were no statistically significant differences. As stated above, we also found that only in the IG a few instances of content features, that is, instances in which the learners noticed ideas or expressions, were reported. In our view, perhaps noticing was reported less frequently in the pairs because they were able to resolve some of their problems together as they discussed what to write and, therefore, they felt that it was not necessary to write this noticing down (as shown in Table 6, there was only one FN related to content in the CollG vs. nine in the IG). In further research, it

would be necessary to analyze the students' conversations to see how the noticing they discuss corresponds (or not) to the noticing they report in the written form. In contrast, the individuals did not have the opportunity to discuss with a peer, and perhaps that led them to take more notes.

6.3 Rewriting stage: incorporations and text quality

Our results showed a moderate percentage of incorporations in both groups (44.26%). As was the case with the amount of noticing, our results also yielded lower rates than other studies. Hanaoka (2007) with adults and Coyle and Roca de Larios (2014) with children reported a rate over 90% for their model groups. When looking at whether the incorporations had been previously reported, most came from the FNs at Stage 2 or from both stages. Thus, the model texts seemed to be an effective trigger for meaningful and solvable noticing. In fact, we only found two incorporations of features that were noticed at Stage 1 exclusively, and both of them corresponded to features solvable in the model texts.

Finally, the pairs incorporated 72.72% of the features correctly (8 out of 11) while the individuals incorporated only 47.37% (9 out of 19). This difference, however, was not statistically significant and, therefore, both groups also seemed to be similar in this respect. The amount of noticing in the current study was low, and these results could only be tentative but, overall, even after only one treatment and without using their notes to write the second composition, young children were able to incorporate 60% of the features correctly. This positive finding concurs with the study conducted by Cánovas Guirao et al. (2015), in which the acceptability of the incorporations in a model group reached a rate of 64%. This suggests that the accuracy of the incorporations will be an interesting aspect to analyze in future studies. Such analysis will help us to fully understand the potential of model texts not only to notice but also to correctly incorporate the features in students' subsequent writings and, ultimately, in the students' interlanguages.

The differences between Draft 1 and Draft 2 in terms of CAF were not statistically significant. By contrast, the holistic ratings showed a significant improvement from Draft 1 to Draft 2 in both groups; this is in line with the findings of Coyle and Roca de Larios (2014). Although the small sample size and shortness of the compositions preclude us from making any strong claims, it seems that the CAF measures have failed to grasp the global improvements that our students made in their drafts. This finding highlights the importance of combining quantitative and holistic measurements as well as the need to refine CAF measures and to become aware of their limitations (Housen and Kuiken 2009). As Norris and Ortega (2009) suggest, while it is advisable to use CAF measures to enable comparison with previous research, it is also advisable to supplement them with other measures

that are suitable for a specific study. In line with this, some recent studies with adult learners have examined how the connection between CAF measures and holistic scores contributes to a deeper understanding of the potential of writing processes (Choi and Deane 2021). In our study, using a holistic analysis helped us to identify some global improvements that could have remained unnoticed if we had only used CAF measures.

7 Conclusions

The present study investigated the use of model texts with YLs of English. More specifically, we wanted to examine the noticing that was triggered in the writing process (PFNs) and in the process of comparing students' drafts with a model text (FNs). The features learners were able to incorporate, their correction, and the possibility that the second draft may be better than the previous one were also part of our analysis. All this was addressed by analyzing the production of students working in pairs and students working individually.

Our findings, with no significant differences between the IG and the CollG, showed few instances of noticing while writing and also few, but more, instances of noticing while comparing with the model texts. Thus, our students were able to notice and incorporate features, although in smaller quantities than participants in previous research. Perhaps the model texts used in other studies were better suited to the students' needs (Hanaoka 2007), or perhaps our students noticed less due to their young age and low level or, as some authors have suggested, due to their lack of training (Coyle et al. 2018; García Mayo and Loidi Labandibar 2017; Yang and Zhang 2010).

As for the types of noticing, in line with all previous studies, lexical features occupied most of the learners' attention and only when comparing with the model texts did other types of features arise. Again, this points to the importance of implementing model texts in a more guided manner and extended over several attempts to help learners benefit from a broader range of features. To do this, teachers could guide students when comparing their drafts to model texts by showing them examples of the different categories of features they could notice. Also, model texts could be accompanied by noticing tables including a variety of categories (lexical, grammatical, and discourse) (García Mayo and Loidi Labandibar 2017), and students could be asked to annotate at least one feature of each category.

Our students were also able to incorporate features into their final drafts although, as was the case with the noticed features, the rates were lower than those of previous studies. This could also be attributed to their age, level, and lack of training; we also need to bear in mind the fact that our students wrote the final draft anew, without having a copy of the original one or of their own previously reported PFNs and FNs

(unlike Coyle and Roca de Larios 2014; Hanaoka 2007; Martínez Esteban and Roca de Larios 2010). In spite of this, when correction was considered, our students incorporated 60% of the noticed features correctly. This suggests that the learners did benefit from the model texts and also reinforces previous findings reporting high rates of acceptability in the students' incorporations (Cánovas Guriao et al. 2015). Likewise, while the limited quantity of noticing and incorporations did not provide the necessary impetus to make the second draft better in terms of CAF measures, a better holistic grade was obtained for the second draft in both groups. This holistic improvement adds to our understanding of the impact of model texts on text quality, supports the findings in Coyle and Roca de Larios (2014), and shows the importance of including both quantitative and qualitative measures in order to gauge improvements in draft quality.

Our study has some limitations that need to be acknowledged but that also open some lines for further research. The sample of participants was relatively small, as was the quantity of noticing reported by them. The small sample and small quantity of noticing could affect the statistical results obtained; not finding an effect might be due to little power and not to lack of treatment effects (Larson-Hall 2010).

Another limitation to bear in mind is the fact that our task was very new to the students, who were not used to writing collaboratively or using model texts, and who had only received two training sessions. Future research could include training sessions over longer periods of time (Coyle et al. 2018). In addition, our students only wrote one composition. It would be interesting to analyze more than one instance of writing and also to use similar but different picture prompts to write the final drafts to see if students are able to transfer the noticed features when faced with a new narrative. Also, the model texts were written by two native speakers but had not been validated through previous testing.

Regarding the pair modality, an analysis of the students' conversations (Cánovas Guirao et al. 2015; Yang and Zhang 2010) would be vital to investigate whether the pairs noticed a larger number of features that they discussed but did not report in the written form. Likewise, another line of further research could be to tap into the thoughts of individuals (perhaps by asking them to speak aloud during the task or through the use of think-aloud protocols). The efforts of the individual writers to process feedback must have taken place silently through mental activity (Ellis et al. 1994), and we do not know if they noticed a greater number of features that they did not report.

Finally, future research could ask students about their opinions on the model texts, as these opinions could affect their willingness to incorporate feedback into their own writing. Similarly, measures of task motivation would help provide a deeper understanding of students' outcomes.

In sum, with no significant differences between the IG and the CollG, the children in the present study noticed features mainly when comparing, and these were mainly lexical. Also, the children were able to accurately incorporate a reasonable number of

them, and the written product, when analyzed holistically, improved after they had compared their drafts to the model texts. This leads us to conclude that the impact of model texts on noticing among these students has been limited when compared to other studies but has still been positive for both groups. Therefore, we encourage teachers to introduce model texts in language lessons. Providing feedback via model texts is feasible and, compared to the usual provision of corrections of isolated errors, it promotes a more comprehensive focus on writing as well as a student-generated type of noticing. In light of our results, we also recommend accompanying model texts with training and follow-up sessions in order to promote a greater variety and quantity of noticed features and a greater quantity of (accurate) incorporations, which would, in turn, result in drafts of greater quality.

Appendix 1: Stage 1. Composing

SHEET 1 – PICTURE PROMPT

NAME(S):   DATE: SCHOOL: CLASS:

WRITE 8–10 LINES DESCRIBING WHAT HAPPENS IN THE FOLLOWING PICTURES

_____
_____
_____
_____
_____
_____
_____
_____
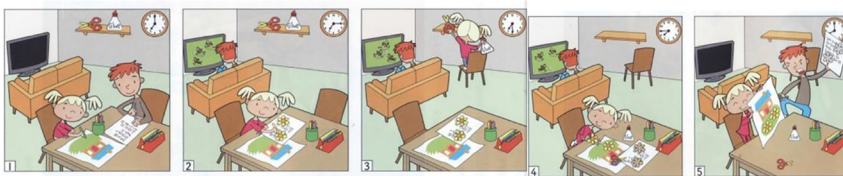
Write down the problems that you have had when writing the composition. You can do so in English, Spanish or Basque. You can use the back of the page too.

Examples: "I didn't know how to say "X" and I said …"; "I'm not sure if the sentence "X" is correct because…."

Appendix 2: Stage 2. Comparison

SHEET 2 – MODEL TEXTS
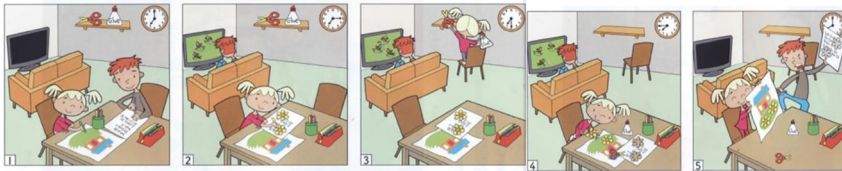
NAME(S):   **DATE:**   **SCHOOL:**   **CLASS:**

Compare your composition to the model texts below

Source. From Cambridge English. (2014). Young Learners. Young Learners English Tests
(YLE). Sample Papers. Flyers. Practice Test 3. Cambridge University Press, p. 3.

MODEL A

It's Monday evening and Tom is doing his homework while his sister Katie does some drawing. After a while, Tom gets bored and decides to watch the football on TV. Katie has drawn a lovely house with a garden but she wants to put some flowers on the grass. She sees Tom's homework and starts to draw flowers on it.

Soon Katie's flowers are finished so she gets some scissors and glue, cuts them out and sticks them onto her picture. Proud of her work, Katie shows her picture to her brother. When Tom realises where the flowers have come from and sees his ruined homework, he is furious!

MODEL B

Tom and Katie are brother and sister and they are in the sitting room of their house. They are sitting at a table. It is 7 o'clock. Tom is doing his homework and Katie is drawing a picture of a house and garden. At 7.15 Tom is watching TV and Katie is drawing another picture of two big flowers. She is drawing the flowers on Tom's piece of paper! At 7.30 she gets the glue and the scissors. At 7.45 she has cut out the two flowers and is sticking them both on the picture of her garden. At 8 O'clock Tom sees what Katie has done and is very angry but she is very happy with her picture.

Write down the differences between your composition and the model texts and explain if you have learnt something with the model texts. You can do so in English, Spanish or Basque. You can use the back of the page too.

Appendix 3: Qualitative measurement

Guidelines to global evaluation of writing adapted from Storch (2005).

The writing is assessed on a score out of 5. This score evaluates the writing mainly in terms of structure and task fulfilment. In order to fulfil the task, the writing needs to include the description of the main elements that appear on the pictures and the narration of what happens should also be clear.

1. This is a very well written text. It is well structured. It contains a clear and complete description of the pictures and the narration of the story is logical. Ideas are clearly organized and good use is made of linking words/phrases.
2. This is a good text. The text has a clear overall structure. All pictures are described and the narration of the story is easy to follow most of the time. Ideas are generally well organized and linking words/phrases are generally used appropriately.
3. This is a satisfactory text. It has an overall structure, but the description of some pictures may be incomplete and the narration of the story hard to follow. Linking words/phrases may be missing or used inappropriately.
4. This is an adequate text. The text is difficult to follow because the description is very incomplete and the narration is not well organized. There is a general lack of linking words/phrases. There might be repetitions.
5. This is a poorly written text. It is poorly organized and difficult to follow. Description and narration are poor or absent.

Appendix 4: Individual results for noticing and incorporations

	Individual					Pairs			
Participant	PFNs	FNs	Incorporations	Correct	Participant	PFNs	FNs	Incorporations	Correct
1	0	3	3	3	P1	1	4	0	0
2	3	8	4	3	P2	2	5	2	1
3	3	7	3	0	P3	3	6	2	0
4	0	6	1	0	P4	1	6	1	1
5	3	3	0	0	P5	1	3	2	2
6	1	6	2	0	P6	0	1	1	1
7	2	9	2	1	P7	1	1	0	0
8	0	5	1	1	P8	2	3	0	0
9	4	3	2	0	P9	1	8	1	1
10	0	5	1	1	P10	1	1	1	1
11	3	5	0	0	P11	1	2	1	1
Total	**19**	**60**	**19**	**9**	**Total**	**14**	**40**	**11**	**8**

References

Bitchener, John. 2016. To what extent has the published written CF research aided our understanding of its potential for L2 development. *International Journal of Applied Linguistics* 167(2). 111–131.

Cánovas Guirao, Josefa, Julio Roca de Larios & Yvette Coyle. 2015. The use of models as a written feedback technique with young EFL learners. *System* 52(1). 63–77.

Choi, Ikkyu & Deane Paul. 2021. Evaluating writing process features in an adult EFL writing assessment context: A keystroke logging study. *Language Assessment Quarterly* 18(2). 107–132.

Cohen, Andrew D. 1983. Reformulating compositions. *TESOL Newsletter* 6(12). 4–5.

Collins, Laura & Carmen Muñoz. 2016. The foreign language classroom: Current perspectives and future considerations. *Modern Language Journal* 100. 133–147.

Coyle, Yvette & Josefa Cánovas Guirao. 2019. Learning to write in a second language: The role of guided interaction in promoting children's noticing from model texts. *CLIL Journal of Innovation and Research in Plurilingual and Pluricultural Education* 2(1). 21–30.

Coyle, Yvette, Josefa Cánovas-Guirao & Julio Roca de Larios. 2018. Identifying the trajectories of young EFL learners across multi-stage writing and feedback processing tasks with model texts. *Journal of Second Language Writing* 42. 25–43.

Coyle, Yvette & Julio Roca de Larios. 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(3). 451–485.

Coyle, Yvette & Julio Roca de Larios. 2020. Exploring young learners' engagement with models as a written corrective technique in EFL and CLIL settings. *System* 95. 102374.

Ellis, Rod. 2009. Task-based language teaching: Sorting out the misunderstandings. *International Journal of Applied Linguistics* 19(3). 221–246.

Ellis, Rod, Yoshiro Tanaka & Asako Yamazaki. 1994. Classroom interaction, comprehension, and the acquisition of L2 word meanings. *Language Learning* 44(3). 449–491.

Fernández Dobao, Ana. 2012. Collaborative writing tasks in the L2 classroom: Comparing group, pair, and individual work. *Journal of Second Language Writing* 21(1). 40–58.

García-Mayo, María del Pilar (ed.). 2021. In , *Working collaboratively in second/foreign language learning*, Vol. 30. Berlin: Mouton de Gruyter.

García-Mayo, María del Pilar & Udane Loidi Labandibar. 2017. The use of models as written corrective feedback in EFL writing. *Annual Review of Applied Linguistics* 37. 110–127.

Hanaoka, Osamu. 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. 459–479.

Hanaoka, Osamu & Shinichi Izumi. 2012. Noticing and uptake: Addressing pre-articulated covert problems in L2 writing. *Journal of Second Language Writing* 21(4). 332–347.

Hidalgo, María Ángeles & Amparo Lázaro-Ibarrola. 2020. Task repetition and collaborative writing by EFL children: Beyond CAF measures. *Studies in Second Language Learning and Teaching* 10(3). 501–522.

Housen, Alex & Folkert Kuiken. 2009. Complexity, accuracy, and fluency in second language acquisition. *Applied linguistics* 30(4). 461–473.

Kang, Eun Young & Han Zhaohong. 2015. The efficacy of written corrective feedback in improving L2 written accuracy: A meta-analysis. *Modern Language Journal* 99. 1–18.

Larson-Hall, Jenifer. 2010. *A guide to doing statistics in second language research using SPSS*. New York: Routledge.

Lázaro-Ibarrola, Amparo & María Ángeles Hidalgo. 2021. Task repetition and collaborative writing with child EFL learners. Language Teaching for Young Learners 3(2). 275–299.

Luquin, María & María del Pilar García-Mayo. 2020. Collaborative writing and feedback: An exploratory study of the potential of models in primary EFL students' writing performance. *Language Teaching for Young Learners* 2(1). 73–100.

Luquin, María & María del Pilar García-Mayo. 2021. Exploring the use of models as a written corrective feedback technique among EFL children. *System* 98. 102465.

Lyster, Roy, Kazuya Saito & Masatoshi Sato. 2013. Oral corrective feedback in second language classrooms. *Language Teaching* 46. 1–40.

Malvern, David, Brian Richards, Ngoni Chipere & Pilar Durán. 2004. *Lexical diversity and language development*. New York: Palgrave Macmillan.

Manchón, Rosa M. 2011. Writing to learn the language: Issues in theory and research. In Rosa M. Manchón (ed.), *Learning-to-write and writing-to-learn in an additional language*, 61–82. Amsterdam: John Benjamins.

Martínez Esteban, Noelia & Julio Roca de Larios. 2010. The use of models as a form of written feedback to secondary school pupils of English. *International Journal of English Studies* 10. 143–170.

McDonough, Kim & César García-Fuentes. 2015. The effect of writing task and task conditions on Colombian EFL learners' language use. *TESL Canada Journal* 32(2). 67–79.

McKee, Gerard, David Malvern & Brian Richards. 2000. Measuring vocabulary diversity using dedicated software. *Literary and Linguistic Computing* 15. 323–337.

Murtiningsih, Sri Rejeki. 2016. Collaborative writing in an EFL context. *Journal of Foreign Language Teaching and Learning* 1(1). 82–90.

Neumann, Heike & Kim McDonough. 2014. Exploring the relationships among student preferences, prewriting tasks, and text quality in an EAP context. *Journal of English for Academic Purposes* 15. 14–26.

Neumann, Heike & Kim McDonough. 2015. Exploring student interaction during collaborative prewriting discussions and its relationship to L2 writing. *Journal of Second Language Writing* 27. 84–104.

Norris, John M. & Lourdes Ortega. 2009. Towards an organic approach to investigating CAF in instructed SLA: The case of complexity. *Applied Linguistics* 30(4). 555–578.

Polio, Charlene. 2012. The relevance of second language acquisition theory to the written error correction debate. *Journal of Second Language Writing* 21. 375–389.

Schmidt, Richard W. 1990. The role of consciousness in second language learning. *Applied Linguistics* 11. 129–158.

Shehadeh, Ali. 2011. Effects and student perceptions of collaborative writing in L2. *Journal of Second Language Writing* 20(4). 286–305.

Storch, Neomy. 2005. Collaborative writing: Product, process, and students' reflections. *Journal of Second Language Writing* 14. 153–173.

Storch, Neomy. 2011. Collaborative writing in L2 contexts: Processes, outcomes, and future directions. *Annual Review of Applied Linguistics* 31. 275–288.

Storch, Neomy (ed.). 2013. *Collaborative writing in L2 classrooms*. Bristol: Multilingual Matters.

Storch, Neomy. 2016. Collaborative writing. In Rosa M. Manchón & Paul Matsuda (eds.), *Handbook of second and foreign language writing*, 387–406. Boston/Berlin: De Gruyter Mouton.

Storch, Neomy. 2019. Written corrective feedback from sociocultural theoretical perspectives: A research agenda. *Language Teaching* 51(2). 262–277.

Storch, Neomy. 2021. Collaborative writing: Promoting languaging among language learners. In María del Pilar & Mayo García (eds.), *Working collaboratively in second/foreign language learning*, 13–34. Boston/Berlin: De Gruyter Mouton.

Swain, Merrill. 1985. Communicative competence: Some roles of comprehensible input and comprehensible output in its development. In Susan Gass & Carolyn Madden (eds.), *Input in second language acquisition*, 235–253. Rowley, MA: Newbury House.

Swain, Merrill. 2000. The output hypothesis and beyond: Mediating acquisition through collaborative dialogue. In James P. Lantolf (ed.), *Sociocultural theory and second language learning*, 97–114. Oxford: Oxford University Press.

Swain, Merril & Sharon Lapkin. 1998. Interaction and second language learning: Two adolescent French immersion students working together. *The Modern Language Journal* 82(3). 320–337.

Swales, John M. & Christine B. Feak. 1994. *Academic writing for graduate students*. Ann Arbor: University of Michigan Press.

VanPatten, Bill. 1990. Attending to form and content in the input: An experiment in consciousness. *Studies in Second Language Acquisition* 12(3). 287–301.

Villarreal, Izaskun & Nora Gil-Sarratea. 2020. The effect of collaborative writing in an EFL secondary setting. *Language Teaching Research* 24(6). 874–897.

Wigglesworth, Gillian & Neomy Storch. 2009. Pair versus individual writing: Effects on fluency, complexity and accuracy. *Language Testing* 26(3). 445–466.

Wigglesworth, Gillian & Neomy Storch. 2012. What role for collaboration in writing and writing feedback. *Journal of Second Language Writing* 21(4). 364–374.

Wong, Wynne. 2001. Modality and attention to meaning and form in the input. *Studies in Second Language Acquisition* 23. 345–368.

Yang, Luxin & Ling Zhang. 2010. Exploring the role of reformulations and a model text in EFL students' writing performance. *Language Teaching Research* 14(4). 464–484.