

*This is an accepted manuscript of an article published by Taylor & Francis in  
The Language Learning Journal on 16 Dec 2022, available online:  
<https://doi.org/10.1080/09571736.2022.2141828>*

**Collaborative Writing among Young EFL Learners in a School Context:  
Product and Process**

Amparo Lázaro-Ibarrola  
ORCID iD 0000-0002-3016-5901  
Institute for Advanced Social Research (I-COMMUNITAS)  
Public Universtiy of Navarre

María Ángeles Hidalgo  
ORCID iD 0000-0002-7131-1880  
Institute for Advanced Social Research (I-COMMUNITAS)  
Public Universtiy of Navarre

**Abstract**

The recent surge in studies on collaborative writing (CW) has provided valuable insights into the product and process of writing. When writing together, adults tend to produce better texts and generate and resolve a large number of language-related episodes (LREs). Also, analyses of the dialogues of collaborative writers show that learners are able to co-construct knowledge and mainly focus their attention on the generation of ideas. As for young learners (YLS), the very few studies comparing jointly and individually written texts have not identified any advantages in the collaboratively written drafts. Besides, while YLS also produce and resolve LREs in CW tasks, no study to date has provided a thorough analysis of their dialogues. To address these gaps, this study compares the products of primary school learners of English as a foreign language (EFL) aged 11–12 writing in pairs (N=20) and individually (N=19) and provides a thorough analysis of pair talk (process). Results suggest that collaborative writers produced more accurate texts and focused most of their efforts on the generation of ideas and on the discussion and successful resolution of LREs. In light of these results the implementation of CW with YLS is encouraged.

**Keywords:** collaborative writing; young learners; CAF; writing-to-learn; language-related episodes.

## 1. Introduction

Until recently, second language acquisition (SLA) researchers and second language teachers and examiners regarded writing as a solitary task (Storch, 2011; Wigglesworth & Storch, 2009). Students only collaborated when producing drafts during brainstorming and peer review activities (Fernández Dobao, 2012; Storch & Wigglesworth, 2007). In the past decade, however, the new perspective on writing tasks offered by the writing-to-learn approach (Manchón, 2011, 2020), together with the well-established advantages of peer work (Kim, 2020; Swain & Lapkin, 2013), have aroused interest in the conjoint potential of writing and peer collaboration (Hirvela & Belcher, 2016; Storch, 2016).

This interest has generated numerous studies implementing collaborative writing (CW) tasks with adult learners (Fernández Dobao, 2012; Storch, 1999, 2005; Teng, 2017, 2021; Teng & Huang, 2021; Wigglesworth & Storch, 2009) and also, in recent years, with young learners (YLS) (see Lázaro-Ibarrola (2023) for a review). In these studies, researchers have tried to identify the effects of collaboration on the writing product, on the one hand, while at the same time trying to understand the writing process, which becomes accessible thanks to the learners' conversations while writing.

Regarding the products of CW, studies with adult learners have demonstrated that jointly written drafts show greater accuracy than those written individually (Fernández Dobao, 2012; Storch & Wigglesworth, 2007; Wigglesworth & Storch, 2009). With YLS, the available studies comparing the products of collaborative and individual writers have not reported any difference between jointly and individually written drafts (Hidalgo & Lázaro-Ibarrola, 2020; Lázaro-Ibarrola, 2021). However, research is still scant.

As for the process of writing, the majority of analyses of students' dialogues while writing have exclusively focused on the parts "in which students talk about the language they are producing, question their language use, or correct themselves or others" (Swain & Lapkin, 1998, p. 326), that is, on language-related episodes (LREs). These analyses seem to indicate that both adults and YLS are able to discuss and successfully resolve many linguistic issues, with adults focusing more on lexis (Fernández Dobao, 2012; Storch & Wigglesworth, 2007; Wigglesworth & Storch, 2009) and YLS focusing more on grammar and mechanics (Hidalgo & García-Mayo, 2021; Hidalgo & Lázaro-Ibarrola, 2020). However, to date, only the studies by Storch (2005) and Wigglesworth and Storch (2009) have offered a thorough analysis, not only of the

LREs but also of the totality of the pair talk (Storch, 2005; Wigglesworth & Storch, 2009; Teng, 2021). Wigglesworth & Storch (2009) identified three phases of writing (planning, composing, and revising) and several types of episodes depending on what the students were talking about (task clarification, idea generation, structure, revision, and LREs). Their analysis of pair talk was, therefore, unique, and added to the literature by helping to describe the process of writing as a whole and in greater depth.

In sum, the conjoint potential of writing and collaboration is gaining ground in SLA research, but some gaps need to be addressed to enhance our understanding of CW tasks, especially in the case of YLs. With this age group we need more studies that include pairs and individuals performing writing tasks, and specifically, studies that analyse students' talk in its entirety. The current study intends to address these two research gaps by exploring the potential of CW among YLs (ages 11–12) of English in a school context. To do so, it compares the written products of a group of students writing individually and a group writing in pairs, and it offers a detailed analysis of pair talk. Thus, we hope to contribute to current understanding of CW and of YLs, a growing but under-researched group in the field of SLA (Collins & Muñoz, 2016) and, in particular, in the field of second language (L2) writing tasks (Lázaro-Ibarrola, 2023).

## **2. Literature Review**

### ***2.1. Collaborative Writing***

In the past decade, theoretical and empirical studies on CW have experienced an unprecedented increase (Lázaro-Ibarrola, 2023; Storch, 2021). CW can be defined as the conjoint production of a single text written by two or more authors (Storch, 2019). This implies that collaboration takes place throughout the entire writing process and that all writers should have control over their production and contribute to decision-making at all levels (content, structure, and language). In Storch's (2013, 2018) words, all writers should co-author and co-own the text.

The implementation of CW activities has been said to facilitate language learning in different ways (García Mayo, 2021). First, and framed within the theoretical underpinning of the output hypothesis (Swain, 1985, 2000), the mere production of language (oral or written) is seen as a powerful learning tool. In the specific case of written output (vs. oral output), and within the 'writing-to-learn' approach, Manchón (2011, 2020) explains that L2 writing activities facilitate L2 learning because they require a slow pace, allow time to reflect on language and demand greater precision and

greater accuracy than oral output. Also, learners writing collaboratively (vs. learners writing individually) appear to benefit from the collaborative dialogue they necessarily engage in while deciding what to write and how. In these dialogues learners are able to pool their resources together, sort out problems and build new knowledge (Storch, 2021). Finally, learners frequently express positive feelings about writing tasks performed in collaboration (Lázaro-Ibarrola & Villarreal, 2021; Calzada & García Mayo, 2021; Storch, 2005, 2013). This suggests that CW might also have a positive impact on students' task motivation which, in turn, could counterweight the inherent challenges of writing in an L2 (Lázaro-Ibarrola, 2023).

However, a characteristic of research on L2 writing is that most studies do not focus on the impact of the collaboration itself; rather, they are implemented with students working in pairs but explore other factors, such as task repetition (Hidalgo & Lázaro-Ibarrola, 2020; Lázaro-Ibarrola & Hidalgo, 2021), amount and functions of first language use (L1) (Antón & DiCamilla, 1998; Swain & Lapkin, 2000, 2001), pre-task modeling (Kim & McDonough, 2011), focus on form (Azkarai & Kopinska, 2020; Calzada & García Mayo, 2020; Kopinska & Azkarai, 2020), and WCF via model texts (Coyle et al., 2018; Coyle & Roca de Larios, 2020; Lázaro-Ibarrola, 2021; Luquin & García Mayo, 2020, 2021; Villarreal & Lázaro-Ibarrola, 2022). Thus, although a great bulk of research has implemented CW tasks, empirical studies comparing the individual and collaborative writers are extremely scarce.

## ***2.2. Studies on CW with adults***

Several studies investigating adult students writing together and individually have demonstrated that learners produce better texts when writing in pairs. Particularly, they tend to produce shorter but better texts (Storch, 2005); texts of greater accuracy (Fernández Dobao, 2012; Teng & Huang, 2021; Wigglesworth & Storch, 2009); and, in some cases, texts with greater lexical and grammatical complexity (Villarreal & Gil-Sarratea, 2019, with adolescents).

In addition to analysing the products of writing, CW allows researchers to get insights into the process of writing by analysing students' oral conversations while writing. In particular, researchers have focused on the learners' production of LREs. In general, the research findings indicate that learners generate and resolve a large number of LREs, mainly focused on lexis and grammar (and, more rarely, on punctuation and spelling). Also, when pairs and groups are compared, groups seem to generate more

LREs (Fernández Dobao, 2012) and when the individual gains of the members of the pair are considered, all students seem to benefit (Villarreal & Gil-Sarratea, 2019).

On the other hand, very few studies have tried to provide an analysis of students' talk in its entirety, that is, an analysis of the dialogues which does not only include the LREs, but which considers the content of the whole conversation (Storch, 2005; Wigglesworth & Storch, 2009). One of the available studies focusing on the analysis of the learners' whole dialogues was conducted by Storch (2005). This author examined the conversations of nine pairs of university students at an Australian university and concluded that, while writing together, the students mainly collaborated when generating ideas, and their collaboration allowed them to give and receive feedback on language (operationalised as LREs), which was not possible for students writing individually. Expanding on these findings, Wigglesworth and Storch (2009) analysed the dialogues of 48 pairs of Australian university students enrolled in post-graduate courses and showed that collaboration afforded learners the opportunity to share ideas and also to discuss and pool their knowledge about language. In fact, the students, on average, devoted 30% of the time to discussion of language issues, which suggests that they viewed this as an important aspect of writing.

### ***2.3. Studies on CW with Young Learners***

As for studies on CW with YLs, some researchers exploring the potential of written corrective feedback and task repetition have reported gains in text quality when students write individually (Roothoof et al., 2022) and when students write in pairs (Villarreal & Lázaro-Ibarrola, 2022), however, very few authors have compared the quality of individually and collaboratively written drafts (Lázaro-Ibarrola, 2021; Lázaro-Ibarrola & Hidalgo, 2021); very few provide an analysis of the LREs students produce while writing; and none has analysed all the aspects and elements of the students' conversations (Lázaro-Ibarrola, 2023). Starting with the few studies addressing the quality of individually and jointly written texts, Lázaro-Ibarrola and Hidalgo (2021) conducted a study comparing individual writers (N=19) and collaborative writers (N=40) learning English as a foreign language (EFL) in primary school (age 11). The learners repeated either the exact same task or a task with the same procedure but different content. Analyses of the students' products revealed gains in terms of text quality in the second draft in the same task repetition group, but, unlike previous

findings with adults (Fernández Dobao, 2012; Storch, 2005; Villarreal & Gil-Sarratea, 2019; Wigglesworth & Storch, 2009), the products written in collaboration were not better (or worse) than those written individually. Likewise, a study by Lázaro-Ibarrola (2021) with YLs analysed (among other aspects) the quality of the texts written by individuals and by pairs who had received written corrective feedback in the form of model texts and did not report any differences.

YLs' talk while repeating a CW task has also been analysed. Hidalgo and García-Mayo (2021) explored the impact of task repetition type on the LREs generated by 40 YLs (age 11–12). They found that most of the LREs generated while writing the two drafts were successfully resolved. Also, they reported that the majority of LREs were form-focused, a category in which these authors included morphosyntax, spelling, and phonology. In line with these findings, Hidalgo and Lázaro-Ibarrola (2020) also explored the effects of repetition on the LREs generated in the conversations of 10 pairs of EFL learners (age 12). Their findings showed that the YLs were able to generate and resolve a large number of LREs while discussing what and how to write, and that most LREs were form-focused.

In sum, in spite of the increasing interest in CW, to date, very few studies have compared compositions produced collaboratively with compositions produced individually by YLs (Lázaro-Ibarrola, 2021; Lázaro-Ibarrola & Hidalgo, 2021). In addition, research offering thorough analyses of the entire process of writing is still scant (Storch, 2005; Wigglesworth & Storch, 2009) and is non-existent among young populations (Lázaro-Ibarrola, 2023). With these research gaps in mind, the current study investigates the product and process of CW by (i) comparing compositions written in pairs and individually, and (ii) analysing the conversations conducted by the pairs while composing. The research questions, accordingly, are as follows:

1. To what extent do YLs' collaboratively and individually written texts exhibit differences in terms of text quality?
2. What is the nature of the conversations that pairs engage in during the writing process?

### **3. Methods**

#### ***3.1. Participants***

This study analysed data from 59 EFL learners (mean age 11.33) in their final year of primary education in a state-funded school in Spain. The participants had been learning

English in the same school since they were 3, and all spoke Spanish as their main language. At the time of data collection, there were three parallel classes in the school, and the study was conducted in these three classes. The school had agreed to participate voluntarily, and informed consents of the participants' parents or guards were obtained before data collection.

With the help of the school teachers, the paper-based reading and writing parts of the Cambridge Key English Test (KET) were selected, as they seemed best able to measure the students' level: the school objective for English at the end of primary school was the A2 level on the Common European Framework of Reference for Languages, and the children were familiar with the test itself. In light of the results, several students were excluded because they had a much lower level of English. In the end, 59 young EFL learners (27 girls and 32 boys) participated in the study. The learners were randomly assigned to two different conditions. A total of 40 learners, 18 girls and 22 boys, were assigned to the collaborative group. These students were distributed into 20 pairs assigned by the school teachers, who tried to pair up students who could work together well. The remaining 19 learners, 9 girls and 10 boys, were assigned to the individual group. Based on the results of the English test, the level of proficiency of both groups (individual and collaborative) was similar, as shown by a Mann-Whitney U test ( $Z = -0.745$ ;  $p = .456$ ).

### ***3.2. Procedure***

The task used in the study was a picture description task, consisting of a comic strip with five pictures showing a little anecdote. The pictures were a sample from the Young Learners English Tests. The learners had to write a composition describing what was happening in the pictures. Two weeks prior to data collection, the students' teacher had conducted a pilot task to familiarise students with CW and with the use of a comic strip as a writing prompt. Additionally, several comic strips had been presented to the teachers, who were asked to select an appropriate one in terms of level and content. One was selected for the pilot task and another for the study.

The study was conducted in one single session, when the students were approaching the end of the school year. The learners writing the composition in pairs sat together and were given two minutes to look at the pictures and speak to each other describing what was happening in the story. Then, they were asked to write the story together. Each dyad had to produce a single text. The pairs were audio- and video-

recorded throughout the process. The learners in the individual writing condition were also given two minutes to look at the pictures and, like the learners working in pairs, after two minutes, they had to write the story. Regarding timing, in line with previous studies comparing collaborative and individual writing (Wigglesworth & Storch, 2009), the collaborative writers were given more time (30 minutes) than the individual writers (20 minutes). None of the participants had difficulties in completing the task within the allotted time. Table 1 summarizes the characteristics and procedure in both groups.

*[Table 1 near here]*

### **3.3. Codification Measures**

#### *3.3.1. Analysis of the Compositions.*

The compositions were analysed quantitatively, by means of complexity, accuracy and fluency (CAF) measures and, also, qualitatively by means of a holistic rubric (see Appendix A).

Quantitative analyses were conducted using CAF measures because these are the most widely used measures of quality in SLA writing (Housen & Kuiken, 2009). Our choice of measures for each of the CAF components was based on the elements of analyses employed by authors in previous studies on CW with adults (Storch, 2005; Wigglesworth & Storch, 2009) and YLs (Lázaro-Ibarrola, 2021). Table 2 summarises the quantitative measures that were used to analyse the writing produced by the participants (see Appendix B for further notes on codification).

*[Table 2 near here]*

As Table 2 shows, fluency was measured by considering the average number of words per composition, the proportion of T-units per text, and the average number of clauses per text. Complexity comprised structural (grammatical) and lexical features. Structural complexity was measured as number of clauses per T-unit and the number of dependent clauses divided by the total number of clauses (Foster & Skehan, 1999). Lexical complexity was measured by means of lexical density (D), which considers the number of lexical words divided by the total number of words (Malvern & Richards, 2000; Malvern et al., 2004) and has the advantage of being less sensitive to text length (MacWhinney, 2000). Accuracy was measured calculating the proportion of error-free



T-units divided by the total number of T-units, and the proportion of error-free clauses divided by the total number of clauses. Both proportions were expressed as percentages (Fernández Dobao, 2012; Storch, 2005; Storch & Wigglesworth, 2007; Wigglesworth & Storch, 2009). Lexical errors (word choice) were only considered when the word used obscured meaning (Fernández Dobao, 2012; Wigglesworth & Storch, 2009).

In addition to the above, and following the advice from some authors advocating a combination of measures of complexity, accuracy, and fluency (CAF) and measures of global assessment (Choi & Deane, 2020; Norris & Ortega, 2009), we also performed a qualitative evaluation of the texts. A five-scale global evaluation scheme used in previous research with the same picture prompt (Hidalgo & Lázaro-Ibarrola, 2020; Lázaro-Ibarrola, 2021; Lázaro-Ibarrola & Hidalgo, 2021) and developed from the one proposed by Storch (2005) was used (Appendix A). The goal of this rubric was to measure the content, structure, and degree of task fulfilment.

### 3.3.2. *Analysis of Pair Dialogues.*

In line with previous studies (Storch, 2005; Teng, 2021; Wigglesworth and Storch, 2009) we analysed our learners' discourse in detail. Specifically, our units of analysis were based on Storch (2005) and Wigglesworth and Storch (2009). Thus, the dialogues were analysed at three different levels:

- Level 1: Time spent on planning, composing and revising.
- Level 2: Identification of episodes and their focus.
- Level 3: Analysis of LREs.

Level 1 offers a broad division of the time spent to complete the task into three main stages: before writing (planning); while writing (composing); and after writing (revising). Thus, the time spent on planning corresponds to the time before students start to write. The time spent on composing comprises all the time the students devote to completing their draft. Finally, the time spent on revising corresponds to the time students devote to revising all or part of the composition once it has been written. The episodes (level 2) are parts of the learners' conversations, which consist of one or several turns, and which have a specific focus on one of the following activities, illustrated in Table 3 with examples from our database.

*[Table 3 near here]*

At Level 3, all the transcripts were analysed to identify LREs. These episodes constitute a key element in providing insight into learners' understanding of the language itself (Wigglesworth & Storch, 2009).

LREs were further divided into lexis-focused (L-LREs), in which students discuss the meaning of a word, search for a new word, or choose between different words, as in (1); form-focused (F-LREs), where students discuss grammatical issues (including tense, adverbs, possessive 's, and pronouns) as in (2); and episodes related to the mechanics of language (M-LREs), in which the children focus on spelling or punctuation, as in (3). All examples come from our database.

(1) Lexis-focused LRE

\*CHI1: of a house or a house eh with the (...) sabes cómo se dice techo? (*do you know how to say roof?*)

\*CHI2: no.

\*CHI1: eh (...) with the door red.

(2) Form-focused LRE

\*CHI1: took the chair and step on the chair to took.

\*CHI2: step on a chair to take.

\*CHI1: step on a chair.

\*CHI2: to take the glue and the scissors.

\*CHI1: to take the glue.

(3) Mechanics LRE

\*CHI1: when the football match finish.

\*CHI1: when the football. Two "1", another "1".

Following Swain (1985) and García Mayo and Azkarai (2016), and with the aim of gaining a deeper understanding of the potential of LREs, we further categorised them in terms of their outcome, that is, whether they were correctly resolved (see 2 and 3 above) or not (see 1).

### **3.4. Inter-rater Reliability**

The participants' written products (39 texts) were coded by one of the authors of this paper, an experienced English teacher and SLA researcher. A random sample of drafts (17%) was also coded by a research assistant. Inter-rater agreement was calculated using Cohen's Kappa. There was almost perfect agreement between the two researchers

for complexity ( $\kappa = .935$ ;  $p < .001$ ), accuracy ( $\kappa = .833$ ;  $p < .001$ ), and fluency ( $\kappa = .954$ ;  $p < .001$ ). In the holistic analysis, given that subjectivity could play a greater role, a greater number of drafts were analysed by the second rater (38% of the drafts), and substantial agreement was found ( $\kappa = 0.749$ ,  $p < .001$ ).

Following Wigglesworth and Storch (2009), a sample of the conversations (50%) entertained by the pairs was also analysed by the same researcher. Inter-rater agreement was checked; perfect agreement was found between the two researchers for phases ( $\kappa = 1.000$ ;  $p < .001$ ), and substantial agreement was observed for episodes ( $\kappa = .605$ ;  $p < .001$ ) and LREs ( $\kappa = .676$ ;  $p < .001$ ).

Given the normal distribution of our data, Independent Samples t-tests were used to check for the statistical significance of the results. The significance level was set at  $\alpha = 0.05$ .

## 4. Results

### 4.1. Analysis of the Compositions (CAF and Holistic Measures)

Table 4 shows the results for measures of fluency.

*[Table 4 near here]*

As Table 4 shows, the number of words, clauses and T-units per text are strikingly similar for the compositions written individually and collaboratively. The compositions written individually were not longer than those written collaboratively. The average length of the compositions written individually was 105.32 words (SD = 19.39), whereas compositions written in pairs contained 109.85 words on average (SD = 21.94). Similar results were obtained for the rest of the fluency measures: Individual texts contained an average of 11.47 T-units (SD = 2.59) and 14.37 clauses per text (SD = 3); and collaborative compositions had 11.7 T-units (SD = 2.36) and 15 clauses (SD = 3.14). These small differences between the two groups were not statistically significant (words per text  $t(37) = -.68$ ,  $p = .499$ ; T-units per text:  $t(37) = -.28$ ,  $p = .777$ ; clauses per text:  $t(37) = -.64$ ,  $p = .526$ ). Thus, these results suggest that collaboration had no impact on fluency.

Likewise, our data and the corresponding statistical analysis conducted for structural (Table 5) and lexical complexity (Table 6) showed that the differences between the compositions written in pairs and those written individually did not reach

statistical significance. That is, the compositions of the two groups displayed similar structural (clauses per T-unit:  $t(37) = -.32, p = .744$ ; percentage of dependent clauses per total clauses:  $t(37) = -.45, p = .655$ ) and lexical complexity (number of words:  $t(37) = -.68, p = .499$ ; lexical diversity:  $t(37) = .83, p = .412$ ).

*[Table 5 near here]*

*[Table 6 near here]*

Table 7 shows that texts composed by pairs seem to be more accurate. The percentage of error-free T-units in the collaboratively written texts was 54.27% over the total number of T-units, whereas this percentage was 38.53% in the individual compositions. This difference is, in fact, statistically significant ( $t(37) = -2.37, p = .023$ ). The percentage of error-free clauses was also higher in the compositions written in pairs (54.66%) than in those written individually (45.52%), although the differences did not reach statistical significance ( $t(37) = -1.03, p = .308$ ).

*[Table 7 near here]*

Finally, Table 8 presents the results for the holistic ratings. The individuals obtained a slightly higher mean score on their compositions. However, this difference was not statistically significant ( $t(37) = 1.07, p = .289$ ), which suggests that collaboration did not influence this aspect.

*[Table 8 near here]*

#### **4.2. Analysis of the Pair Dialogues (Phases and Episodes)**

Table 9 illustrates the results regarding the time spent on the different phases of writing (planning, composing, and revising).

*[Table 9 near here]*

As shown in Table 9, most of the time was spent on composing (around 65.8%, which corresponds to approximately 10:43 minutes on average), followed by planning, which

took up 25.3% of the time, that is, about 3:33 minutes on average. On the contrary, most students barely paid attention to the revision stage, which only amounts to 8.9% of the time (1:27 minutes approximately). This trend was followed homogeneously by all the pairs (see Appendix C for distribution of time per pair) with the only exception of Pair 6, who devoted around the same amount of time to planning and revising (2:09 vs. 2:44).

Next, Table 10 summarises the proportion of turns devoted to the different categories of discussion codified as episodes. The distribution of episodes per pair can be found in Appendix D.

*[Table 10 near here]*

As the table shows, most of the turns (61.4%) were devoted to generating ideas and discussing the content of the compositions. These discussions mainly revolved around describing what the participants saw in the pictures and interpreting what was happening in the story. The second most common episode type found in our data set was the discussion of linguistic issues in the form of LREs (25.41% of the turns were devoted to these discussions). The remaining episodes were given little attention by the learners. Task clarification episodes mainly happened at the beginning of the task (in the planning stage) and represented 6.62% of the turns. In them, the learners read the instructions and decided how to deal with the task mainly by discussing who was going to write. Revision activities were also very scarce (taking up only 3.38% of the turns), and all of them consisted of learners re-reading their text aloud. Finally, episodes dealing with text structure and the organization of ideas were the least frequent and represented 2.48% of the total amount of turns.

Except for the episodes related to the structure and organization, which were not present in the output of three pairs, and the two dyads that did not revise their compositions, there were instances of all episodes in all pairs and in similar proportions. All pairs generated LREs, and content discussion was the main focus of their conversations. The main difference lay in the proportion of turns devoted to these two categories, with some pairs having almost the same number of turns in both types of episodes (e.g., Pair 1, content episodes: 38.76%; LREs: 30.89%), but with most having a higher percentage of turns in the episodes devoted to content (the most extreme being Pair 5, content episodes: 82.24%; LREs: 10.28%).

Table 11 presents the results obtained for LREs.

*[Table 11 near here]*

As shown in Table 11, the number of LREs varied from pair to pair (the minimum being 5 and the maximum 17), but all pairs produced instances of them. Thus, the episodes devoted to discussing the language (LREs) represented 34.42% of the total number of episodes identified in the oral production of our participants, which corresponded to 25.41% of the total number of turns.

Mechanical episodes were the most common type, representing 42.1% of the total LREs identified. Almost all of these episodes dealt with spelling issues. In fact, only five episodes dealt with punctuation, and all were generated by the same pair. Spelling seemed to be important for our learners, with spelling issues being the most common focus of the LREs (35 LREs out of a total of 95), generated by all pairs except one.

Form-focused LREs were the second most frequent type (35.78%). All pairs engaged in at least one form-focused LRE and, when looking into the issues under discussion (see Table 12), they were limited to four aspects (tense, adverbs, Saxon genitive, and pronouns), with verb tenses concentrating most of the attention (22 LREs out of 34).

*[Table 12 near here]*

Finally, the percentage of lexical LREs was also remarkable (22.1%). Except for four episodes dealing with word choice, the participants in the present study mainly concentrated on discussing the meaning of a given word. There were also four pairs who did not produce any instances of lexical LREs (see Table 13).

*[Table 13 near here]*

Finally, Table 14 summarises the results regarding the rate of correctly resolved LREs.

*[Table 14 near here]*

The percentage of successfully resolved LREs is quite high, especially in the case of form-focused (70.58%) and mechanical LREs (69.4%). However, if we look at lexical LREs, the percentage is much lower than in the other two types (38.1%).

## **5. Discussion and Conclusion**

The present study has investigated the potential of CW among YLs, comparing individually and jointly written drafts and adopting a holistic approach to analyse the process of writing. To this end, the drafts of 40 children working in pairs were compared to the drafts of 19 children working individually, and 50% of the conversations entertained by the pairs during the process of writing were analysed in depth.

When analysing the products, and in line with research on adult populations (Fernández Dobao, 2012; Storch & Wigglesworth, 2007; Teng & Huang, 2021; Wigglesworth & Storch, 2009) but in contrast with the few studies with YLs (Hidalgo & Lázaro-Ibarrola, 2020; Lázaro-Ibarrola, 2021; Lázaro-Ibarrola & Hidalgo, 2021), we found that the collaborative writings of young EFL learners were more accurate than those written individually. Fluency and complexity did not present any differences; this contrasts with Villarreal and Gil-Sarratea (2019), who also found advantages in these aspects. Regarding the holistic analyses, results were similar in both conditions (Storch, 2005). Thus, we could say that our findings only partly align with the few studies on CW with YLs. These studies did not identify any differences related to individual vs. collaborative conditions in either holistic or CAF measures when comparing drafts after model texts (Lázaro-Ibarrola, 2021) and after task repetition (Lázaro-Ibarrola & Hidalgo, 2021). The present study has found advantages in accuracy but not in holistic ratings. This difference is hard to interpret, but it is relevant because it hints at potential advantages for accuracy in collaboratively written products with YLs upon mere task repetition. Interestingly, our results fully coincide with Wigglesworth and Storch (2009), who also found statistical significance in error-free T-units but not in error-free clauses, which suggests that, as in our study, the gains in accuracy existed but were modest. This could also be the reason why no differences in holistic ratings were found. In other words, the superiority identified in accuracy might not be powerful enough to give an overall impression of better quality.

The analyses of the conversations during the writing process provided us with positive evidence that could contribute to reinforcing previous findings (Storch, 2005;

Wigglesworth & Storch, 2009). First, and in line with previous research (Storch, 2005; Wigglesworth & Storch, 2009), participants were able to collaborate on the different stages of writing, interacting when planning their compositions and, above all, when generating ideas. On the negative side, learners seemed to pay little attention to the revision of their written products once they had finished the task. This information should be taken into account by language teachers when carrying out writing tasks. If the revision stage is encouraged by practitioners, the quality of the final product might improve. Also positive was the fact that the YLs were able to generate a great number of LREs (25.41% of the total number of episodes). When looking into the different types, mechanics (42.1%) and form-focused LREs (36.8%) accounted for most of these episodes, while lexical LREs (22.6%) were less prevalent. Our findings contrast with Wigglesworth and Storch (2009), who found that 55.50% of the LREs were focused on lexical aspects, but they are in line with other studies with YLs that have reported a focus on grammar and mechanics (Hidalgo & García-Mayo, 2021; Lázaro-Ibarrola & Hidalgo, 2021). Perhaps, in the present study, the vocabulary generated by the picture prompt was quite easy, and the children did not need to discuss vocabulary items so much (García Mayo & Azkarai, 2016), or perhaps the maturity of adults allows them to deal with lexical aspects (find synonyms, consider register, or look for a more precise term), while children either know or do not know a given word. In line with this idea, another positive finding was the important rates of successful resolution of the LREs (Hidalgo & García-Mayo, 2021; Hidalgo & Lázaro-Ibarrola, 2020). The learners were able to successfully resolve around 70% of the linguistic difficulties encountered regarding grammar and spelling, and around 40% regarding vocabulary. Thus, the conversations were very effective in resolving grammar and spelling issues although they did not seem to contribute so much to resolving lexical problems. In sum, the analysis of the LREs suggests that our participants were mainly focused on two aspects: spelling and grammar. This highlights how aware children seem to be of the importance of correction, which might be a reflection of the form-focused approach to teaching often adopted in foreign language contexts. In addition, we noticed how YLs devote much of the task time to discussing the content (in terms of turn-taking), and their focus on language aspects, albeit manifest, remains secondary. When LREs are the only object of analysis, we do not get a clear picture of what percentage of the learners' attention they represent. The analysis of the rest of episodes that take place during CW helps to do this. LREs are indeed very relevant, but so are other aspects.



The present study has several limitations. For instance, it would be interesting to examine the individual gains that each member of the pair obtains from collaboration (Villarreal & Gil-Sarratea, 2019). Also, we are aware that the relatively small group sizes and the brevity of the compositions constitute a limitation (Lázaro-Ibarrola, 2023; Storch, 2005) and might affect the statistical results (Larson-Hall, 2010). Therefore, these results need to be viewed with caution, and more research with larger groups of students needs to be conducted to confirm (or challenge) the superiority in accuracy that we identified in our group of YLs writing collaboratively. Additionally, tasks that require texts of a longer nature may also provide different results for fluency. Adopting a broader perspective, the benefits of technology in enhancing collaborative writing, which have been demonstrated for adults (Teng, 2021), could also be tested with YLs. This would constitute a timely line of research of great pedagogical value for current EFL lessons in primary schools.

In sum, while further research is clearly needed, we believe that CW has been positive for our participants, not only because the collaborative writers generated a slightly more accurate draft, but also because they were able to interact in order to co-write a composition. These YLs did so in the foreign language, discussing the content and ideas, focusing on language, and successfully sorting out a great number of linguistic problems. All this leads us to encourage the introduction of CW tasks among language learners in primary schools. On the one hand, students' drafts could be of greater quality and, on the other hand, students could enjoy the benefits of both oral and written tasks. During CW, learners have the opportunity to produce (oral and written) output in the target language, to interact with their peers, and to co-construct new knowledge.

### **Acknowledgements**

This study was supported by the Spanish Ministry of Economy and Competitiveness, National Research Agency and European Regional Development Fund-AEI/FEDER/EU [FFI2016-74950-P]. We would also like to thank the schools, the teachers and children for their willingness and generosity to participate in this project.

## References

- Antón, M., & DiCamilla, F. (1998). Socio-cognitive functions of L1 collaborative interaction in the L2 classroom. *Canadian Modern Language Review*, 54(3), 314–342. <https://doi.org/10.3138/cmlr.54.3.314>
- Azkarai, A., & Kopinska, M. (2020). Young EFL learners and collaborative writing: A study on patterns of interaction, engagement in LREs, and task motivation. *System*, 94(1). <https://doi.org/10.1016/j.system.2020.102338>
- Calzada, A., & García Mayo, M. P. (2020). Child EFL grammar learning through a collaborative writing task. *Language in Language Learning and Teaching*, 19–40. <https://doi.org/10.1075/llt.55.01cal>
- Calzada, A., & García Mayo, M. P. (2021). Child learners' reflections about EFL grammar in a collaborative writing task: When form is not at odds with communication. *Language Awareness*, 30(1), 1–16. <https://doi.org/10.1080/09658416.2020.1751178>
- Choi, I., & Deane, P. (2020). Evaluating writing process features in an adult EFL writing assessment context: A keystroke logging study. *Language Assessment Quarterly*, 1 (26), 107–132. <https://doi.org/10.1080/15434303.2020.1804913>
- Collins, L., & Muñoz, C. (2016). The foreign language classroom: Current perspectives and future considerations. *The Modern Language Journal*, 100, 133–147. <https://doi.org/10.1111/modl.12305>
- Coyle, Y., Cánovas-Guirao, J., & Roca de Larios, J. (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. <https://doi.org/10.1016/j.jslw.2018.09.002>
- Coyle, Y., & Roca de Larios, J. (2020). Exploring young learners' engagement with models as a written corrective technique in EFL and CLIL settings. *System*, 95, 1–14. <https://doi.org/10.1016/j.system.2020.102374>
- Fernández Dobao, A. (2012). Collaborative writing tasks in the L2 classroom: Comparing group, pair, and individual work. *Journal of Second Language Writing*, 21(1), 40–58. <https://doi.org/10.1016/j.jslw.2011.12.002>
- Foster, P., & Skehan, P. (1999). The influence of sources of planning and focus of planning on task-based performance. *Language Teaching Research*, 3, 215–247.

- García Mayo, M. P. (Ed.). (2021). *Working Collaboratively in Second/Foreign Language Learning* (Vol. 30). Mouton de Gruyter.
- García Mayo, M. P., & Azkarai, A. (2016). EFL task-based interaction. Does task modality impact on language-related episodes? In M. Sato & S. Ballinger (Eds.), *Peer interaction and second language learning: Pedagogical potential and research agenda* (pp. 241–266). John Benjamins.
- Hidalgo, M. Á., & García Mayo, M. P. (2021). The influence of task repetition type on young EFL learners' attention to form. *Language Teaching Research*, 3, 948–952. <http://doi.org/10.1177/1362168819865559>
- Hidalgo, M. Á., & Lázaro-Ibarrola, A. (2020). Task repetition and collaborative writing by EFL children: Beyond CAF measures. *Studies in Second Language Learning and Teaching*, 10(3), 501–522. <https://doi.org/10.14746/ssllt.2020.10.3.5>
- Hirvela, A., & Belcher, D. (2016). Reading/writing and speaking/writing connections: The advantages of multimodal pedagogy. In R. Manchón & P. K. Matsuda (Eds.), *Handbook of Second and Foreign Language Writing* (pp. 587–612). De Gruyter Mouton.
- Housen, A., & Kuiken, F. (2009). Complexity, accuracy, and fluency in second language acquisition. *Applied linguistics*, 30(4), 461–473.
- Kim, Y. (2020). Willingness to engage: The importance of what learners bring to pair work. *Language Awareness*, 29(2), 134–154. <https://doi.org/10.1080/09658416.2020.1743712>
- Kim, Y., & McDonough, K. (2011). Using pretask modelling to encourage collaborative learning opportunities. *Language Teaching Research*, 15(2), 183–199. <https://doi.org/10.1177/1362168810388711>
- Kopinska, M., & Azkarai, A. (2020). Exploring young EFL learners' motivation: Individual versus pair work on dictogloss tasks. *Studies in Second Language Learning and Teaching*, 10(3), 607–630. <https://doi.org/10.14746/ssllt.2020.10.3.10>
- Larson-Hall, J. (2010). *A Guide to Doing Statistics in Second Language Research Using SPSS*. Routledge.
- Lázaro-Ibarrola, A. (2021). Model texts in collaborative and individual writing among EFL children: noticing, incorporations, and draft quality. *International Review of Applied Linguistics in Language Teaching*. <https://doi.org/10.1515/iral-2020-0160>

- Lázaro-Ibarrola, A. (2023). *Child L2 Writers: A Room of their Own*. John Benjamins.
- Lázaro-Ibarrola, A., & Hidalgo, M. A. (2021). Give me a second chance: Task repetition and collaborative writing with child EFL learners. *Language Teaching for Young Learners*, 3(2), 275–299. <https://doi.org/10.1075/ltyl.20009.laz>
- Lázaro-Ibarrola, A., & Villarreal, I. (2021). Are EFL writers motivated or demotivated by model texts and task repetition? Evidence from young collaborative writers. *International Journal of English Studies*, 21(2), 29–55. <https://doi.org/10.6018/ijes.466401>
- Luquin, M., & García Mayo, M. P. (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, M., & García Mayo, M. P. (2021). Exploring the use of models as a written corrective feedback technique among EFL children. *System*, 98, 102465. <https://doi.org/10.1016/j.system.2021.102465>
- MacWhinney, B. (2000). *The CHILDES project: Tools for Analyzing Talk*. 3rd edition. Mahwah, NJ: Lawrence Erlbaum
- Malvern, D., & Richards, B.J. (2000). Validation of a new measure of lexical diversity. In Beers, M., van den Bogaerde, B., Bol, G., de Jong, J. & C. Rooijmans (Eds.), *From Sound to Sentence: Studies on First Language Acquisition* (pp. 81–96). John Benjamins.
- Malvern, D., Richards, B., Chipere, N., & Durán, P. (2004). *Lexical Diversity and Language Development. Quantification and Assessment*. Palgrave Macmillan UK.
- Manchón, R. M. (Ed.). (2011). *Learning-to-write and Writing-to-learn in an Additional Language*. John Benjamins.
- Manchón, R. M. (Ed.). (2020). *Writing and Language Learning: Advancing Research Agendas*. John Benjamins.
- Norris, J., & Ortega, L. (2009). Towards an organic approach to investigating CAF in instructed SLA: The case of complexity. *Applied Linguistics*, 30(4), 555–578. <https://doi.org/10.1093/applin/amp044>
- Roothoof, H., Lázaro-Ibarrola, A., & Bulté, B. (2022). Task repetition and corrective feedback via models and direct corrections among young EFL writers: Draft quality and task motivation. *Language Teaching Research*, 0(0). <https://doi.org/10.1177/13621688221082041>

- Storch, N. (1999). Are two heads better than one? Pair work and grammatical accuracy. *System*, 27(3), 363–374. [https://doi.org/10.1016/S0346-251X\(99\)00031-7](https://doi.org/10.1016/S0346-251X(99)00031-7)
- Storch, N. (2005). Collaborative writing: Product, process, and students' reflections. *Journal of Second Language Writing*, 14(3), 153–173. <https://doi.org/10.1016/j.jslw.2005.05.002>
- Storch, N. (2011). Collaborative writing in L2 contexts: Processes, outcomes, and future directions. *Annual Review of Applied Linguistics*, 31, 275–288. <https://doi.org/10.1017/S0267190511000079>
- Storch, N. (2013). *Collaborative writing in L2 classrooms*. Multilingual Matters.
- Storch, N. (2016). Collaborative writing. In R. M. Manchón & P. Matsuda (Eds.), *Handbook of second and foreign language writing* (pp. 387–406). De Gruyter Mouton.
- Storch, N. (2018). Written corrective feedback from sociocultural theoretical perspectives: A research agenda. *Language Teaching*, 51(2), 262–277. <https://doi.org/10.1017/S0261444818000034>
- Storch, N. (2019). Collaborative writing. *Language Teaching*, 52, 40–59. <https://doi.org/10.1017/S0261444818000320>
- Storch, N. (2021). Collaborative writing: Promoting languaging among language learners. In M. P. García Mayo (Ed.), *Working collaboratively in second/foreign language learning* (pp. 13–34). De Gruyter Mouton.
- Storch, N., & Wigglesworth, G. (2007). Writing tasks: Comparing individual and collaborative writing. In M. P. García Mayo (Ed.), *Investigating tasks in formal language learning* (pp. 157–177). Multilingual Matters.
- Swain, M. (1985). Communicative competence: Some roles of comprehensible input and comprehensible output in its development. In S. M. Gass & C. Madden (Eds.), *Input in Second Language Acquisition* (pp. 235–253). Newbury House.
- Swain, M. (2000). The output hypothesis and beyond: Mediating acquisition through collaborative dialogue. In J. Lantolf (Ed.), *Sociocultural Approaches to Second Language Research* (pp. 97–115). Oxford University Press.
- Swain, M., & Lapkin, S. (1998). Interaction and second language learning: Two adolescent French immersion students working together. *Modern Language Journal*, 82(3), 320–337. <https://doi.org/10.1111/j.1540-4781.1998.tb01209.x>

- Swain, M., & Lapkin, S. (2000). Task-based second language learning: The uses of the first language. *Language Teaching Research*, 4(3), 251–274.  
<https://doi.org/10.1177%2F136216880000400304>
- Swain, M., & Lapkin, S. (2001). *Focus on Form through Collaborative Dialogue: Exploring Task Effects*. Longman, Pearson Education.
- Swain, M., & Lapkin, S. (2013). A Vygotskian sociocultural perspective on immersion education: The L1/L2 debate. *Journal of Immersion and Content-Based Language Education*, 1, 101–129. <https://doi.org/10.1075/jicb.1.1.05swa>
- Teng, M. F. (2017). The effectiveness of group, pair and individual output tasks on learning phrasal verbs. *The Language Learning Journal*, 48(2), 1–14.  
<https://doi.org/10.1080/09571736.2017.1373841>
- Teng, M. F. (2021) Interactive-whiteboard-technology-supported collaborative writing: Writing achievement, metacognitive activities, and co-regulation patterns. *System*, 97, 102426. <https://doi.org/10.1016/j.system.2020.102426>
- Teng, M. F. & Huang, J. (2021). The effects of incorporating metacognitive strategies instruction into collaborative writing on writing complexity, accuracy, and fluency. *Asia Pacific Journal of Education*, 1-20.  
<https://doi.org/10.1080/02188791.2021.1982675>
- Villarreal, I., & Gil-Sarratea, N. (2019). The effect of collaborative writing in an EFL secondary setting. *Language Teaching Research*, 24(6), 874–897.  
<https://doi.org/10.1177/1362168819829017>
- Villarreal & Lázaro-Ibarrola (2022). Models in collaborative writing among CLIL learners of English in primary school: Linguistic outcomes and motivation matters. *System*, 110, 102922. <https://doi.org/10.1016/j.system.2022.102922>.
- Wigglesworth, G., & Storch, N. (2009). Pair versus individual writing: Effects on fluency, complexity and accuracy. *Language Testing*, 26(3), 445–466.  
<https://doi.org/10.1177%2F0265532209104670>

## **Appendices**

### **Appendix A. 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.

- 5 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 organised and good use is made of linking words/phrases.
- 4 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 organised 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.
- 2 This is an adequate text. The text is difficult to follow because the description is very incomplete and the narration is not well organised. There is a general lack of linking words/phrases. There might be repetitions.
- 1 This is a poorly written text. It is poorly organised and difficult to follow. Description and narration are poor or absent.

## Appendix B. Notes on Codification Measures

T-units are defined as independent clauses and all their attached or embedded dependent clauses (Wigglesworth & Storch, 2009). Given the number of coordinated sentences found in our corpus, it is important to explain that, following Young (1995), co-ordinate clauses were counted as two t-units.

Clauses are classified into the following:

- Independent clauses: grammatical structures that contain a subject and a verb and can stand on their own.
- Dependent clauses: clauses containing a verb that cannot stand on their own. They may be introduced by adverbials, relative clauses, or noun clauses.

T-units appear separated by //

Clauses appear separated by /

Clauses in which the verb was elided were also included, for example, “one day one girl and one boy sitting in a chair next to each other”

Hello my name is Maddie // and I am going to tell you / what I did this afternoon//. I was drowing with my brother Tom,/ when my brother stand up //and swith on the TV//. I was still drowing //and I decided to take the scissor // and I glue the flowers / that I was drowing in Tom’s page and in the grass of my drowing//. When the football match finished / Tom came // and he get anoid with me because of the homework//.



### Appendix C. Time Spent on Phases per Pair

Pairs	Total	Planning	Composing	Revision
1	15:07	3:16	11:50	0
2	14:37	3:31	9:33	1:32
3	13:38	2:43	9:38	1:15
4	14:51	2:30	11:23	0:56
5	16:03	3:36	11:22	1:04
6	12:32	2:09	7:38	2:44
7	12:23	3:15	7:08	1:59
8	22:29	7:00	11:36	3:52
9	21:18	4:10	16:57	0:10
10	14:40	3:20	10:13	1:05
TOTAL	2:43:20	0:41:30	1:47:18	0:14:32
Percentage	100%	25.3%	65.8%	8.9%

### Appendix D. Distribution of Episodes per Pair

Pair	Task clarification		Idea generation & content discussion		Structure		Revision		LREs	
	%	Turns	%	Turns	%	Turns	%	Turns	%	Turns
1	13.66	25	37.7	69	6.13	12	9.29	17	32.79	60
2	6.71	10	73.15	109	0.00	0	0.00	0	20.13	30
3	8.40	10	62.18	74	0.00	0	5.04	6	24.37	29
4	7.39	13	65.91	116	3.06	7	0.57	1	22.16	39
5	5.50	6	80.73	88	0.54	1	0.92	1	11.93	13
6	3.45	4	54.31	63	1.20	2	7.76	9	32.76	38
7	4.72	5	55.66	59	5.74	9	0.00	0	31.13	33
8	11.22	23	42.93	88	2.67	6	7.80	16	35.12	72
9	1.22	3	74.69	183	0.00	0	0.82	2	23.27	57
10	3.16	5	72.78	115	0.89	2	0.63	1	22.15	35
Total	6.64	104	61.56	964	2.56	39	3.38	53	25.93	406
%		6.59%		61.13%		2.47%		3.36%		25.75%