# Task repetition and collaborative writing by EFL children: Beyond CAF measures

## María Ángeles Hidalgo
Universidad Pública de Navarra, Spain
https://orcid.org/0000-0002-7131-1880
*mangeles.hidalgo@unavarra.es*

## Amparo Lázaro-Ibarrola
Universidad Pública de Navarra, Spain
https://orcid.org/0000-0002-3016-5901
*amparo.lazaro@unavarra.es*

## Abstract

Research into the potential of collaborative writing is relatively new. Similarly, task repetition (TR), which has been claimed to be a valuable tool for language learning, has been rarely explored in the context of writing. Therefore, little is known about the potential of combining TR and collaborative writing, and even less if we focus on young learners (YLs), who constitute a generally under-researched population. With these research gaps in mind, the present study examines the compositions of 10 pairs of learners of English as a foreign language (EFL) (aged 12) who write the same text in response to the same picture prompt three times over a three-week period. Our analysis includes the language-related episodes (LREs) that learners generate while writing collaboratively and, also, a thorough analysis of the three drafts that students produce, including quantitative (complexity, accuracy and fluency (CAF)) and holistic measures. Results show that learners' compositions improve with repetition when measured by holistic ratings although CAF measures fail to grasp this improvement. As for the LREs, a great amount was found, most of the episodes were focused on form, most were successfully resolved and their amount declined with TR. In light of these results we argue in favor of the inclusion of holistic measures when analyzing students' productions and discuss the positive effects of collaborative writing in the context of TR with YLs.

*Keywords*: young learners; task repetition; collaborative writing; holistic measures; CAF

## 1. Introduction

The use of pair and group work in language classrooms is anchored in firm pedagogical and theoretical bases (Storch, 2011) and has been frequently investigated in the context of oral language (Mackey & Gass, 2006), whereas the study of collaborative writing is still relatively new (Abrams & Byrd, 2017; Storch, 2011). Likewise, task repetition (TR), which has been claimed to offer students great learning opportunities by allowing them to shift their focus from content to form (Bygate & Samuda, 2005), has been rarely explored in the context of writing (Amiryousefi, 2016), where findings from research on oral data can hardly be applicable due to the important differences between the oral and written mode (Gilabert, Manchón, & Vasylets, 2016; Manchón, 2014; Tavakoli, 2014). As TR and pair work have been explored independently, little is known about the potential of combining them, and even less about their potential in the case of writing tasks. Finally, most of the existing literature has focused on adult learners, disregarding a population that is increasing all over the world: young language learners (Collins & Muñoz, 2016; Copland, Garton, & Burns, 2014; Enever, 2018; Pinter, 2017).

In order to shed some light into this research gap, this study examines the compositions of 10 pairs of 12 year-old learners of English who had to write the same narrative in response to a picture prompt three times over a three week period in a classroom context. Our analysis includes the widely used measurement of the main components of linguistic performance (complexity, accuracy and fluency (CAF)) (Housen & Kuiken, 2009; Housen, Kuiken, & Vedder, 2012; Michel, 2017), a holistic assessment of their writings (Storch, 2005), and the analysis of the students' deliberations during the writing process operationalized as language related episodes (LREs) (Swain & Lapkin, 1998). Our findings will help to better understand the potential of task repetition in the context of writing with young learners (YLs).

## 2. Literature review

### 2.1. Collaborative writing

Collaborative writing has been defined as "the production of a text by two or more writers" (Storch, 2016, p. 387). While writing together, the authors are expected to interact, combine their ideas, and co-author and co-own the text, as well as their responsibilities as writers (Ede & Lunsford, 1990; Storch, 2013, 2018). Ideally, in a text written collaboratively, the parts created by each of its authors cannot be identified. Collaborative writing combines the benefits of oral interaction and writing tasks. During interaction, learners engage in meaningful

use of the target language (TL), have opportunities to negotiate for meaning, and produce modified output. In addition, they receive peer support as well as immediate feedback, and are able to co-construct new meaning (Loewen & Sato, 2018; Long, 1983; Storch, 2013; Swain, 2005). The benefits of writing are many. One of the advantages is the extra time learners have to pay attention to meaning and form, which is not as available during oral-only tasks (Manchón, 2014; Storch, 2016). Given the lack of spontaneity and immediacy of writing, as well as the access writers have to their production, anxiety might also be lower than in oral communication (Tavakoli, 2014). Moreover, writing has been claimed to encourage the use of language structures that are not normally employed orally (Williams, 2012). Finally, the written modality demands higher levels of accuracy, as errors tend to be less tolerated (Schoonen, Snellings, Stevenson, & van Gelderen, 2009).

Studies that have compared writing tasks carried out in pairs with tasks completed individually have reported gains in accuracy regarding target words and structures when learners collaborate (Nassaji & Tian, 2010; Storch, 2007; Teng, 2020). In addition, learners writing collaboratively have been reported to produce shorter but better texts in terms of grammatical accuracy, complexity and task fulfilment (Storch, 2005). Learners writing collaboratively have also been found to initiate and solve more LREs than they do when performing oral-only tasks (Adams & Ross-Feldman, 2008; García Mayo & Azkarai, 2016). Also, when asked, learners have expressed positive views towards writing in collaboration with their peers (Storch, 2005).

However, to date, these findings are based on research into adult collaborative writing. Little is known about children writing collaboratively and whether the claims summarized above hold for this specific learner group (Coyle & Roca de Larios, 2014).

## 2.2. Task repetition

The repetition of communicative situations occurs in everyday life. We often need to perform the same tasks and chores more than once in our life. We have to go to the shops, to the bank, or just interact with our neighbors in the lift. TR constitutes, therefore, a common human activity. Bygate (2018) recently defined the construct of TR as "the repetition of a given configuration of purposes, and a set of content information" (p. 2). This definition underlines the idea that nothing can be exactly repeated and that, consequently, changes may happen from one performance to the next. These constitute, in fact, the key elements of TR: how learners' performances vary from one iteration to another, and how these changes relate to language acquisition (Bygate, 2018).

TR influences the way learners perform a task, and the language they use to deal with it. TR has been found to help learners to produce improved output (Bygate, 1996, 2001; Lambert, Kormos, & Minn, 2017; Sample & Michel, 2014). By repeating a task, learners' attention is diverted from conceptualizing the meaning they want to convey during the first iteration, to the formulation of their message in subsequent encounters with the task (Bui, Ahmadian, & Hunter, 2018; Bygate, 1996). Most research has addressed the effect of TR on adult learners' oral performance, and scarce attention has been paid to the potential of TR for YLs (Pinter, 2006, 2007, 2011).

In any case, research to date both on adult and child populations concurs that gains have always been found with TR although there are differences regarding the aspects that show greater improvements. In general, fluency gains have been reported, whereas the evidence regarding complexity and accuracy is more variable (Ahmadian & Tavakoli, 2011; Bagheri, Rahimi, & Riasati, 2012; Bret Blasco, 2014; Bygate, 2001; Bygate & Samuda, 2005; García Mayo, Imaz Agirre, & Azkarai, 2017; Hidalgo, 2018; Hu, 2018; Lynch & Maclean, 2000; Pinter, 2006, 2007, 2011; Sample & Michel, 2014). The fact that findings regarding some aspects are inconclusive (Bui et al., 2018; Bygate, 2018) may be partly due to the great diversity of variables analyzed (context, age, level, tasks, and time span between repetitions) (Lázaro-Ibarrola & Hidalgo, 2017).

It is also important to make a distinction between *same TR*, the most widely explored type, in which learners repeat the exact same task, and *task-type repetition* (procedural repetition), in which students repeat the same task type but with different content (Kim, 2013; Kim & Tracy-Ventura, 2013; Payant & Reagan, 2018). With oral data from junior high school Korean students, Kim (2013), and Kim and Tracy-Ventura (2013) compared these two types of repetition and found that learners' interest and focus on form (measured by their use of LREs) decreased when repeating the same task in comparison to learners who repeated different versions of the same task type. However, they do not recommend any method over the other, since no significant differences were found between the groups. Payant and Reagan's (2018) study also showed that LREs decreased with exact TR and that learners focused mainly on the meaning of the message they want to convey, producing more meaning-focused LREs. On the other hand, these authors suggest that exact TR had greater benefits as regards the production of LREs. Finally, they reported that most LREs were correctly solved.

Despite the body of work addressing TR in relation to different aspects of language performance, only a few studies have analyzed the effect of TR on writing (Amiryousefi, 2016; Manchón, 2014; Nitta & Baba, 2014). One of the few studies addressing TR and written performance is Amiryousefi's (2016). This author analyzed the effects of exact TR and procedural TR on low-proficiency EFL

learners' (mean age 23.56) computer-mediated individual written production. His results provided positive evidence of the benefits of both TR types for writing, although some differences were found. The compositions by the exact TR group improved significantly in terms of fluency (measured as numbers of words, clauses and T-units) and in one of his accuracy measures (the percentage of error-free clauses), whereas the procedural TR group only improved in two of the fluency measures (namely number of words and clauses per text). Nitta and Baba (2014) explored the effect of these two types of TR on writing over time. In their longitudinal study, they found that procedural TR had a marked effect on lexical and grammatical aspects, whereas the influence of exact TR was limited. Nevertheless, they suggest that the benefits of TR may be more noticeable in the long term.

In a very similar context to that of our study, Hidalgo and García Mayo (2019) examined the effect of TR on the production of LREs by YLs while performing a collaborative writing task. Contrary to most research to date, their participants initiated more form-focused than meaning-focused LREs. On the other hand, they also reported that most LREs were correctly solved and that LREs decreased significantly with exact TR.

## 3. The study

### 3.1. Research questions

The present study analyzes the effects of exact TR on the collaborative writing of 10 pairs of EFL learners. Our first aim is to find out if learners are able to generate better texts (measured quantitatively and holistically) with TR. Also, we want to understand how TR affects the LREs that learners generate while writing, that is, whether it affects the amount, the type or their ability to successfully resolve them. Therefore, our research questions are the following:

1. How do learners' drafts change (quantitatively and qualitatively) with TR?
2. How does TR affect the number, nature and resolution of learners' LREs?

On the basis of the literature review, our learners' drafts will be expected to improve with repetition, however, it is not clear what specific components might improve more. On the other hand, LREs will be expected to decrease with the repetitions, and will probably be mainly form-focused and correctly solved (Hidalgo & García Mayo, 2019).

## 3.2. Participants and setting

The participants in the present study were 20 EFL learners (mean age 11.39) who attended a Content and Language Integrated Learning (CLIL) program at a state school in the north of Spain. At the moment of data collection their command of the TL was described as an A2 level of the *Common European Framework of References for Languages* (CEFR), as attested by the Cambridge Key English Test (KET) and by school-internal tests.

In the school, the learners followed a CLIL program and their exposure to the TL was approximately 14 hours per week. English language as such was allotted five sessions per week, and the rest of hours of exposure included other subject matters taught through English, such as math, science, art and physical education. This CLIL program was mandatory for all pupils. This eliminates the risk that only the most motivated learners, or those with a higher-than-average command of the TL, would participate in the study.

## 3.3. Procedure

The participants had to work in pairs to write a narrative in response to a picture prompt three times over a three-week period in a classroom context. The pairs were established by the researchers and learners' own teacher, taking into account their personal relationship (to avoid conflict) and, at the same time, trying to make pairs of very similar levels of proficiency. The prompt consisted of a six-picture comic strip (Cambridge English, 2014, p. 3). The dyads sat together and were given two minutes to look at the pictures and speak about them. After the two minutes, they were asked to collaborate to compose the story in writing, with a pen, on a piece of paper. Each dyad had to produce a single composition at each data collection time. The time limit set for students to perform the task was fifteen minutes. The dyads remained the same throughout the experiment.

The participants' deliberations were video and audio recorded and their oral production (30 transcripts, 8 hours approximately) was transcribed into the CHAT (Codes for the Human Analysis of Transcripts) format. Their attention to form, operationalized as LREs (Swain & Lapkin, 1998), was coded using the CLAN (Computerized Language Analysis) tools (MacWhinney, 2000).

## 3.4. Coding and analysis

Our analysis of the learners' written compositions consisted of both quantitative and holistic measures. In both cases we compared the production at Time 1 (henceforth T1) versus the production at Time 2 (T2) and the production at T1 and T2 vs. the production at Time 3 (T3).

L2 performance has been defined as multicomponential in nature and its principal components have been successfully captured in the notions of complexity, fluency and accuracy (CAF) (Housen & Kuiken, 2009; Housen et al., 2012; Michel, 2017). Although there is some controversy regarding how these constructs are operationalized (particularly with fluency and complexity), the three components still are the most reliable tool to measure proficiency (Housen et al., 2012).

Our choice of the specific CAF measures was based on the main measurements used in some previous studies that seemed to be applicable to our data. Thus, complexity was measured in terms of the proportion of dependent clauses and clauses to T-units (Foster, Tonkyn, & Wigglesworth, 2000). T-units are defined as "one main clause plus whatever subordinate clauses happen to be attached to or embedded within it" (Hunt, 1966, p. 735). Also, our measurement of complexity included lexical diversity, which was measured in terms of the type/token ratio (TTR), that is, the number of different words in a text divided by the total number of words (Malvern, Richards, Chipere, & Durán, 2004). For the analysis of accuracy, the least controversial of the three constructs, the percentage of the error-free clauses over the total number of clauses and the number of errors per total number of words were considered (Storch, 2005; Storch & Wigglesworth, 2007; Wigglesworth & Storch, 2009). Finally, fluency was measured in terms of the number of words, clauses and T-units per text (Wolfe-Quintero, Inagaki, & Kim, 1998).

In addition to this, we also took into account the functional dimension of our students' production by carrying out a holistic assessment of their writings (De Jong, Steinel, Florijn, Schoonen, & Hulstijn, 2012; Kuiken, Vedder, & Gilabert, 2010; Pallotti, 2009). While there is no agreement to date as to how functional adequacy is to be defined or assessed (Iwashita, Brown, McNamara, & O'Hagan, 2008), its inclusion is vital in order to obtain a more comprehensive assessment of students' production. In this paper, functional adequacy is measured using Storch's (2005) 5-scale global evaluation scheme, which we adapted to the content of the task we employed. This evaluation considered the content and structure of the text, as well as the degree of task fulfillment (the appendix).

Finally, our study also analyzed the LREs generated in the students' oral interactions during the process of writing their texts. Following previous research in EFL settings (García Mayo & Azkarai, 2016; Hidalgo & García Mayo, 2019; López-Serrano, Roca de Larios, & Manchón, 2019), the LREs were classified according to their linguistic focus, whether they were meaning-focused or form-focused (deliberation over morpho-syntactic aspects, spelling and pronunciation), and to their outcome (resolved or not resolved). Finally, resolved LREs were further classified as target-like, or non-target-like. The codification of LREs is illustrated with examples (Examples 1, 2, 3 and 4) from our own dataset.

(1) Form-focused and target-like resolution.

Student 1: I am going to tell you what I do this afternoon.
Student 2: but we can change instead of saying I am going to tell you what I do…
Student 1: I think that is good no?
Student 2: . . . what I did.
Student 1: what I did this afternoon sorry.

In Example 1, the learners focus on the tense of the verb they want to use. Student 1 starts narrating the story and Student 2 interrupts her to suggest that they should use a different tense, namely, the past tense. Student 1 agrees with her partner and they settle on the past tense. Thus, this LRE has been coded as form-focused with a target-like resolution.

(2) Meaning-focused, form-focused, and target-like resolution.

Student 1: they saw the TV very content.
Student 2: very
Student 1: very content.
Student 2: no . . . he, he, very happy with two . . . with double p.

In Example 2, Student 2 is not satisfied with the term employed by Student 1 and proposes a more target-like word (*happy*). Additionally, she focuses on the spelling of the word (*with double p*). The LREs in this example have been coded as one meaning-focused and one form-focused LRE, both target-like resolved.

(3) Form-focused and non-target-like resolution.

Student 1: the girl . . . put.
Student 2: the flower.
Student 1: puted no? en pasado [*past tense*].
Student 2: ok.
Student 1: because once upon a time there was a boy and a girl.
Student 2: yes in past. In past.
Student 1: puted. The girl puted.

Example 3 represents an instance in which the learners were not able to successfully solve a form-focused LRE. Student 2 seems to think the verb *to put* has a regular past form, so he adds the *-ed* ending. He provides evidence for his decision by going back to the beginning of their composition and emphasizing the past tenses they had used. His partner agrees, and they use a wrong form (*puted**) in their text. This LRE has been coded as form-focused with a non-tar-get-like resolution.

(4) Form-focused and not resolved.

Student 1: my sister at . . .
Student 2: at his painting and also in his . . .
Student 1: . . . at the painting.
Student 2: pero cómo se dice sin querer? [*but how do you say "unintentionally"?*]
Student 1: but . . . but she . . . put . . . pero podemos decir de otra forma [*but we can say it some other way*] . . . but she put.

Example 4 represents an occasion in which the participants do not solve a meaning-focused LRE. Apparently, none of the learners is able to provide the term they want to use (*unintentionally*), and they decide to write something different.

## 3.5. Inter-rater reliability

The participants' written production was coded by one of the authors of this paper. An independent research assistant also analyzed the production of 5 pairs at the three testing times (50% of the data). Both raters held several meetings prior to data coding to agree on their understanding of the measures of analysis and also after their coding in order to solve the few discrepancies on a case-by-case basis. Inter-rater reliability was checked for all measures and the differences between the two raters were very small. Total agreement was reached by the two researchers for the codification of the LREs. Regarding CAF, total agreement was found for complexity and fluency while the greatest number of discrepancies was found in the case of accuracy (93.5% agreement). The holistic ratings for the three compositions reached a global agreement of 92%.

## 3.6. Statistical analysis

As for the statistical analysis, dependent samples *t*-tests were used for data that presented a normal distribution and Wilcoxon Signed-Rank Tests were used for the data that were not normally distributed. The significance level was set at $\alpha$ = .05.

## 4. Results

The results obtained from the analyses of CAF reveal that TR does not seem to have a great influence on any of these three dimensions in the compositions written by the young participants in the present study. Table 1 shows the results for our complexity measures.

Table 1 Complexity measures

| | T1 | T2 | T3 | T1 vs. T2 | | | T2 vs. T3 | | | T1 vs. T3 | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | t | df | p | t | df | p | t | df | p |
| Clauses to T-units | 3.08 | 2.58 | 2.62 | -1.96 | 9 | .08 | .18 | 9 | .85 | 1.36 | 9 | .2 |
| Dependent clauses | 64.07 | 60.25 | 60.05 | -1 | 9 | .33 | -.06 | 9 | .946 | .91 | 9 | .384 |
| Lexical diversity | 51.96 | 54.18 | 52.78 | .95 | 9 | .364 | -.53 | 9 | .603 | -.32 | 9 | .756 |

As illustrated in Table 1, both the proportion of clauses to T-units and the percentage of dependent clauses appear to follow a decreasing tendency. Lexical diversity, on the other hand, seems to increase in the second repetition and decrease again in the third one. Nevertheless, the differences across tasks did not reach statistical significance for any of the different complexity measures. Table 2 features the results from the analysis of the accuracy measures.

Table 2 Accuracy measures

| | T1 | T2 | T3 | T1 vs. T2 | | | T2 vs. T3 | | | T1 vs. T3 | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | t | df | p | t | df | p | t | df | p |
| Error-free clauses over total number of clauses | .39 | .44 | .40 | 1.11 | 9 | .293 | -.68 | 9 | .511 | -.16 | 9 | .876 |
| Errors per total number of words | .12 | .11 | .12 | -.75 | 9 | .469 | 1.37 | 9 | .201 | .17 | 9 | .865 |

As Table 2 shows, there seems to be a slight increase in the percentage of error-free clauses from the first to the second and third compositions, which might hint at an improvement in terms of accuracy. However, as in the case of complexity, there are no statistically significant changes.

Table 3 Fluency measures

| | T1 | T2 | T3 | T1 vs. T2 | | | T2 vs. T3 | | | T1 vs. T3 | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | t | df | P | t | df | p | t | df | p |
| Number of words per text | 113.4 | 112.4 | 115.5 | -.14 | 9 | .886 | .48 | 9 | .638 | -.33 | 9 | .744 |
| Clauses per text | 16.1 | 16.9 | 16.3 | .84 | 9 | .423 | -.55 | 9 | .591 | -.14 | 9 | .890 |
| T-units per text | 5.4 | 6.8 | 6.6 | 3.28 | 9 | .010* | -.3 | 9 | .770 | -1.65 | 9 | .133 |

As for fluency, Table 3 shows the results. The mean number of words, clauses and T-units per composition show a trend to increase with TR. Nevertheless, only the difference in the proportion of T-units between T1 and T2 is significantly different. All the other aspects did not show statistically significant differences.

The holistic analysis, on the other hand, revealed more encouraging results, as can be seen in Table 4. As can be seen, the mean rate obtained in the three drafts improves with TR. The scores of all participants ranged from 2 to

4.5 and all dyads' last composition was the best rated. A statistical analysis shows that the improvement from task to task of the global evaluation of the texts was statistically significant (T1 vs. T2: $Z$ = -2.56, $p$ = .010; T2 vs. T3: $Z$ = -2.07, $p$ = .038; T1 vs. T3: $Z$ = -2.97, $p$ = .003).

Table 4 Holistic ratings

|  |  | Holistic rating |  |
| --- | --- | --- | --- |
| Composition | T1 | T2 | T3 |
| Average rating | 3 | 3.65 | 4 |
| *SD* | .62 | .7 | .57 |

Finally, the analysis of the LREs identified in the pair dialogues shows that interaction related to language was recurrent in all dyads' oral production while writing their texts. Table 5 shows the amount of LREs produced by the ten pairs in each composition. We can see that the discussions of language aspects, operationalized as LREs, seems to decrease with each TR. In fact, a statistical analysis shows that this difference is statistically significant when comparing the first composition to the last one (T1 vs. T3: $Z$ = -2.60, $p$ = .009).

Table 5 LREs produced by the ten pairs

| LREs | T1 | T2 | T3 |
| --- | --- | --- | --- |
| Total | 125 | 95 | 75 |
| Mean | 12.5 | 9.5 | 7.5 |
| *SD* | 4.03 | 4.99 | 5.17 |

Next, LREs were classified as either meaning- or form-focused. Table 6 summarizes the distribution of the LREs in terms of the total turns for each LRE type.

Table 6 LRE types

|  | T1 |  | T2 |  | T3 |  |
| --- | --- | --- | --- | --- | --- | --- |
| Focus | Meaning | Form | Meaning | Form | Meaning | Form |
| Total | 41 | 84 | 25 | 70 | 15 | 60 |
| Percentage | 32.8 | 67.2 | 26.32 | 73.68 | 20 | 80 |

Table 6 clearly shows that form-focused LREs made up the greatest proportion of the total LREs at the three data collection times although there was also a large number of meaning-focused LREs. The difference between the frequency of these two types was statistically significant in the three tasks (T1: $Z$ = -2.20, $p$ = .028; T2: $Z$ = -2.24, $p$ = .025; T3: $Z$ = -2.49, $p$ = .012). As for the effect of TR on the nature of the LREs, the frequency of the percentage of use of meaning-focused LREs decreases significantly from the first task performance to the last

(T1 vs. T3: $Z$ = -2.40, $p$ = .016). On the other hand, the increase in the amount of the percentage of form-focused LREs at T3 is not statistically significant (T1 vs. T3: $Z$ = -1.68, $p$ = .091).

Table 7 Resolution of LREs

|  | T1 | | | T2 | | | T3 | |
|  | Resolved | | Not resolved | Resolved | | Not resolved | Resolved | | Not resolved |
|  | Target | Non-target | | Target | Non-target | | Target | Non-target | |
| Total | 97 | 21 | 6 | 67 | 17 | 10 | 51 | 18 | 6 |
| Percentage | 77.6% | 16.8% | 4.8% | 70.5% | 17.9% | 10.5% | 68% | 24% | 7.9% |

Finally, we addressed the impact of TR on the outcome of the LREs. The results are presented in Table 7. The most relevant finding is that most LREs were target-like resolved. On the other hand, the percentage of the correctly solved LREs appears to follow a decreasing trend, however, this decrease did not reach statistical significance (T1 vs. T2: $Z$ = -1.12, $p$ = .26; T2 vs. T3: $Z$ = -0.35, $p$ = .72; T1 vs. T3: $Z$ = -1.36, $p$ = .17).

## 5. Discussion

The present study has examined the effect of TR on the collaboratively written texts of ten pairs of young EFL learners. More specifically, the two students in each pair worked together while writing the exact same composition three times over a three-week period. Our analysis included the quantitative and holistic analysis of these three compositions as well as the analysis of the quantity, type and resolution of the LREs generated by the learners while writing.

Our first research question addressed the effect of TR on these YLs' written compositions in terms of CAF and holistic ratings. Regarding CAF measures, our results reveal mainly non-significant differences, with only an increase in lexical diversity and in the proportion of T-units at T2. However, the raw numbers seem to suggest a tendency towards a greater number of error-free units, greater lexical diversity and greater fluency in either the second or the third composition. As Storch (2005) suggests, the lack of statistical significance may have to do with the small sample size analyzed in the present study (10 dyads, 20 learners), and the relatively short texts these YLs wrote (113.4 words on average). On the other hand, the holistic ratings help us to complete these results. Each time the learners performed the task the mean score improved significantly. This positive finding is in line with the trends hinted at in the analyses of the CAF measures, which, even though when examined separately did not reach statistical significance, seem to be strong enough to give a better global impression.

Thus, our findings support previous research from the oral domain addressing YLs that report better overall performance across TR (Pinter, 2007; Sample & Michel, 2014) and also suggest that CAF measures are not always able to fully grasp the improvements that students make in their writings.

Our second research question focused on the impact of TR on the quantity, nature and outcome of the LREs YLs initiate while composing a text collaboratively. In our students' production, the overall number of LREs has decreased significantly over time. These learners have worked three times with the exact same content and task procedure, and by the last task performance they are so familiar with both that they may not need to resort to metalinguistic discussions so much. Also, by the third TR, YLs may have already solved most of their doubts and language problems from the first iterations, are able to carry this knowledge to the next performance (Hidalgo & García Mayo, 2019; Payant & Reagan, 2018) and, in line with Sample and Michel's (2014) study with oral data, might also be more able to focus their attention on all three CAF dimensions simultaneously and, therefore, to improve their drafts.

In addition to the above, most of the LREs identified in our data were correctly solved at the three data collection times, also mirroring previous findings (García Mayo & Azkarai, 2016; Hidalgo & García Mayo, 2019; Payant & Reagan, 2018). This evidence underlines the benefits of collaborative writing tasks, which offer learners the opportunity to pool their knowledge together and solve language problems correctly. However, as opposed to most previous research, yet concurring with Hidalgo and García Mayo (2019), the majority of the LREs produced by the YLs in the present study were categorized as form-focused. This seemingly contradictory finding may be related to different factors. First, most previous research studies have addressed adult learners whereas, like Hidalgo and García Mayo (2019), we have worked with primary school learners. Besides, most studies employed oral tasks, whereas we have examined learners' oral interactions while producing a written text. Finally, different categorizations have been employed, which, for instance, consider pronunciation- and spelling-related LREs as lexical-based (Payant & Reagan, 2018). In the current study, on the other hand, we have followed García Mayo and Azkarai (2016), who include the discussion of these aspects in the form-focused category. Our findings regarding the nature of the LREs are more in line with the evidence reported by these authors, who also found that their participants initiated significantly more form-focused LREs when carrying out a written task.

## 6. Conclusion

This study has provided some evidence in favor of the use of collaborative writing and TR with YLs. The repetition of the same composition three times has

helped learners to generate better compositions and to discuss and successfully resolve a great number of LREs, mainly regarding formal aspects, but also with an important number of episodes focused on meaning. Our results also highlight the importance of including functional adequacy among the analytic measures of CAF (Housen et al., 2012). As we reported, CAF measures seemed to show trends of improvement but these did not reach statistical significance. On the contrary, the global assessment revealed that the compositions did, in fact, statistically improve in terms of content, structure and task fulfilment. Therefore, we advocate for the combination of quantitative and holistic measures to obtain a more thorough analysis of students' productions.

From a pedagogical perspective, our study can also offer important implications for teachers of young language learners. Even though the value of pair and group work is well recognized in second language acquisition research, and widespread in education (Storch, 2011), its use in writing lessons is still quite limited (Storch, 2005). With our study we have shed more light on the benefits of peer collaboration during the writing process. Thanks to the LREs the participants initiate, and correctly solve, they are able to successfully complete the tasks. As for the value of TR, our results illustrate how subsequent task performances of the same task lead to improved versions of the original manuscripts.

Certainly, there are some limitations to our research that need to be acknowledged and that in turn open up lines for further research. Studies with a larger sample size and that require the production of longer compositions would be necessary. In addition, research that includes more detailed analysis of the pair dialogues examining other processes learners engage in while carrying out the tasks would also help us to better understand the nature of peer-peer collaboration (López-Serrano et al., 2019). Following Wigglesworth and Storch (2009), these processes would include planning, composition, and revision, as well as the focus of these processes (e.g., task management, generation of ideas, text structure). Another interesting line of research would be the comparison of collaboratively written texts and oral narratives also produced collaboratively.

Despite the limitations to this study, we can conclude that collaborative writing and same TR seem to be beneficial for YLs. Writing together has provided them with opportunities to use the TL in a meaningful context and to share their knowledge on language use and this, combined with the repetition of the same draft, has enabled learners to produce a better final text. Finally, we would like to highlight once more the importance of the inclusion of holistic analyses of students' productions since, as we have seen, they reveal information that otherwise might remain unnoticed.

## Acknowledgements

References

Abrams, Z., & Byrd, D. R. (2017). The effects of meaning-focused pre-tasks on beginning-level L2 writing in German: An exploratory study. *Language Teaching Research*, *21*(4), 434-453. http://doi.org/10.1177/1362168815627383

Adams, R., & Ross-Feldman, L. (2008). Does writing influence learner attention to form? In D. Belcher & A. Hirvela (Eds.), *The oral/literate connection: Perspectives on L2 speaking/writing connections* (pp. 243-267). Ann Arbor, MI: University of Michigan Press.

Ahmadian, M. J., & Tavakoli, M. (2011). The effects of simultaneous use of careful online planning and task repetition on accuracy, complexity, and fluency in EFL learners' oral production. *Language Teaching Research*, *15*(1), 35-59. http://doi.org/10.1177/1362168810383329

Amiryousefi, M. (2016). The differential effects of two types of task repetition on the complexity, accuracy, and fluency in computer-mediated L2 written production: A focus on computer anxiety. *Computer Assisted Language Learning*, *29*(5), 1050-1066. http://doi.org/10.1080/09588221.2016.1170040

Bagheri, M. S., Rahimi, F., & Riasati, M. J. (2012). Communicative interaction in language learning tasks among EFL learners. *Journal of Language Teaching and Research*, *3*(5), 948-952. http://doi.org/10.4304/jltr.3.5.948-952

Bret Blasco, A. (2014). *L2 English young learners' oral production skills in CLIL and EFL settings: A longitudinal study* (Unpublished doctoral dissertation). Universitat Autònoma de Barcelona, Spain.

Bui, G., Ahmadian, M. J., & Hunter, A.-M. (2018). Spacing effects on repeated L2 task performance. *System*, *81*, 1-13. http://doi.org/10.1016/j.system.2018.12.006

Bygate, M. (1996). Effects of task repetition: Appraising the developing language of learners. In J. Willis & D. Willis (Eds.), *Challenge and change in language teaching* (pp. 136-146). Oxford: Macmillan Heinemann.

Bygate, M. (2001). Effects of task repetition on the structure and control of oral language. In M. Bygate, P. Skehan, & M. Swain (Eds.), *Researching pedagogic tasks: Second language learning, teaching, and testing* (pp. 23-48). Harlow: Longman.

Bygate, M. (2018). Introduction. In M. Bygate (Ed.), *Learning language through task repetition* (pp. 1-25). Amsterdam: John Benjamins. http://doi.org/10.1075/tblt.11

Bygate, M., & Samuda, V. (2005). Integrative planning through the use of task repetition. In R. Ellis (Ed.), *Planning and task performance in a second language* (pp. 37-74). Amsterdam: John Benjamins Publishing Company.

Cambridge English. (2014). Young Learners: Young learners English tests (YLE). Sample Papers: Flyers: Practice Test 3. Cambridge: Cambridge University Press.

Collins, L., & Muñoz, C. (2016). The foreign language classroom: Current perspectives and future considerations. *Modern Language Journal, 100*(16), 133-147. http://doi.org/10.1111/modl.12305

Copland, F., Garton, S., & Burns, A. (2014). Challenges in teaching English to young learners: Global perspectives and local realities. *TESOL Quarterly, 48*(4), 738-762. http://doi.org/10.1002/tesq.148

Coyle, Y., & Roca de Larios, J. (2014). Exploring the role played by error correction and models on children's reported noticing and output production in a L2 writing task. *Studies in Second Language Acquisition, 36*(3), 451-485. http://doi.org/10.1017/S0272263113000612

De Jong, N. H., Steinel, M. P., Florijn, A. F., Schoonen, R., & Hulstijn, J. H. (2012). The effect of task complexity on functional adequacy, fluency and lexical diversity in speaking performances of native and non-native speakers. In A. Housen, F. Kuiken, & I. Vedder (Eds.), *Dimensions of L2 performance and proficiency: Complexity, accuracy and fluency in SLA* (pp. 121-142). Amsterdam: John Benjamins Publishing Company.

Ede, L., & Lunsford, A. (1990). *Singular texts/plural authors*. Carbondale: Southern Illinois University Press.

Enever, J. (2018). *Policy and politics in global primary English*. Oxford: Oxford University Press.

Foster, P., Tonkyn, A., & Wigglesworth, G. (2000). Measuring spoken language: A unit for all reasons. *Applied Linguistics, 21*(3), 354-375. http://doi.org/10.1093/applin/21.3.354

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). Amsterdam: John Benjamins.

García Mayo, M. P., Imaz Agirre, A., & Azkarai, A. (2017). Task repetition effects on CAF in EFL child task-based interaction. In M. J. Ahmadian & M. P. García Mayo (Eds.), *Recent perspectives on task-based language learning and teaching* (pp. 11-28). Berlin: De Gruyter Mouton. http://doi.org/10.1515/9781501503399

Gilabert, R., Manchón, R., & Vasylets, O. (2016). Mode in theoretical and empirical TBLT research: Advancing research agendas. *Annual Review of Applied Linguistics, 36*, 117-135. http://doi.org/10.1017/S0267190515000112

Hidalgo, M. Á. (2018). *The oral production and negotiation of meaning of Spanish EFL children in task-supported interaction: The role of age and task repetition* (Unpublished doctoral dissertation). Universidad del País Vasco (UPV/EHU), Spain.

Hidalgo, M. Á., & García Mayo, M. P. (2019). 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

Housen, A., & Kuiken, F. (2009). Complexity, accuracy, and fluency in second language acquisition. *Applied Linguistics*, *30*(4), 461-473. http://doi.org/10.1093/applin/amp048

Housen, A., Kuiken, F., & Vedder, I. (Eds.). (2012). *Dimensions of L2 performance and proficiency: Complexity, accuracy and fluency in SLA* (Vol. 32). Amsterdam: John Benjamins Publishing Company.

Hu, X. (2018). Effects of task type, task-type repetition, and performance criteria on L2 oral production. In M. Bygate (Ed.), *Learning language through task repetition* (pp. 143-169). Amsterdam: John Benjamins Publishing Company.

Hunt, K. (1966). Recent measures in syntactic development. *Elementary English*, *43*, 732-739.

Iwashita, N., Brown, A., McNamara, T., & O'Hagan, S. (2008). Assessed levels of second language speaking proficiency: How distinct? *Applied Linguistics* *29*(1), 24-49. http://doi.org/10.1093/applin/amm017

Kim, Y. (2013). Effects of pretask modeling on attention to form and question development. *TESOL Quarterly*, *47*(1), 8-35. http://doi.org/10.1002/tesq.52

Kim, Y., & Tracy-Ventura, N. (2013). The role of task repetition in L2 performance development: What needs to be repeated during task-based interaction? *System*, *41*, 829-840. http://doi.org/10.1016/j.system.2013.08.005

Kuiken, F., Vedder, I., & Gilabert, R. (2010). Communicative adequacy and linguistic complexity in L2 writing. In I. Bartning, M. Martin, & I. Vedder (Eds.), *Communicative proficiency and linguistic development: Intersections between SLA and language testing research* (pp. 81-100). EUROSLA Monographs Series 1.

Lambert, C., Kormos, J., & Minn, D. (2017). Task repetition and second language speech processing. *Studies in Second Language Acquisition*, *39*(1), 167-196. http://doi.org/10.1017/S0272263116000085

Lázaro-Ibarrola, A., & Hidalgo, M. Á. (2017). Procedural repetition in task-based interaction among young EFL learners. *ITL – International Journal of Applied Linguistics*, *168*(2), 183-202. http://doi.org/10.1075/itl.16024.laz

Loewen, S., & Sato, M. (2018). State of the art article: Interaction and instructed second language acquisition. *Language Teaching*, *51*(3), 285-329. http://doi.org/10.1017/S0261444818000125

Long, M. H. (1983). Linguistic and conversational adjustments to nonnative speakers. *Studies in Second Language Acquisition*, *5*, 177-193. http://doi.org/10.1017/S0272263100004848

López-Serrano, S., Roca de Larios, J, & Manchón, R. M. (2019). Language reflection fostered by individual L2 writing tasks: Developing a theoretically motivated and empirically based coding system. *Studies in Second Language Acquisition*, *41*(3), 503-527. http://doi.org/10.1017/S0272263100004848

Lynch, T., & Maclean, J. (2000). Exploring the benefits of task repetition and recycling for classroom language learning. *Language Teaching Research*, *4*(3), 221-250. http://doi.org/10.1177/136216880000400303

Mackey, A., & Gass, S. (2006). Pushing the methodological boundaries in interaction research: An introduction to the special issue. *Studies in Second Language Acquisition*, *28*(2), 169-178. http://doi.org/10.1017/S0272263106060086

MacWhinney, B. (2000). *The CHILDES project: Tools for analyzing talk* (3rd ed.). Mahwah, NJ: Lawrence Erlbaum Associates.

Malvern, D. D., Richards, B. J., Chipere, N., & Durán, P. (2004). *Lexical diversity and language development: Quantification and assessment*. Hampshire: Palgrave Macmillan.

Manchón, R. M. (2014). The distinctive nature of task repetition in writing: Implications for theory, research, and pedagogy. *Elia*, *14*, 13-41. http://doi.org/10.12795/elia.2014.i14.02

Michel, M. (2017) Complexity, accuracy and fluency (CAF). In S. Loewen & M. Sato (Eds.), *The Routledge handbook of instructed second language acquisition* (pp. 50-68). London: Routledge.

Nassaji, H., & Tian, J. (2010). Collaborative and individual output tasks and their effects on learning English phrasal verbs. *Language Teaching Research*, *14*(4), 397-419. http://doi.org/10.1177/1362168810375364

Nitta, R., & Baba, K. (2014). Task repetition and L2 writing development: A longitudinal study from a dynamic systems perspective. In H. Byrnes & R. M. Manchón (Eds.), *Task-based language learning: Insights from and for L2 writing* (pp. 107-136). Amsterdam: John Benjamins Publishing Company.

Pallotti, G. (2009). CAF: Defining, refining and differentiating constructs. *Applied Linguistics*, *30*, 590-601. http://doi.org/10.1093/applin/amp045

Payant, C., & Reagan, D. (2018). Manipulating task implementation variables with incipient Spanish language learners: A classroom-based study. *Language Teaching Research*, *22*(2), 169-188. http://doi.org/10.1177/1362168816669742

Pinter, A. (2006). Verbal evidence of task related strategies: Child versus adult interactions. *System*, *24*, 615-630. http://doi.org/10.1016/j.system.2006.09.005

Pinter, A. (2007). Some benefits of peer-peer interaction: 10 year-old children practicing with a communicative task. *Language Teaching Research*, *11*(2), 189-207. http://doi.org/10.1177/1362168807074604

Pinter, A. (2011). *Children learning second languages*. London: Palgrave Macmillan.

Pinter, A. (2017). *Teaching young language learners* (2nd ed.) Oxford: Oxford University Press.

Sample, E., & Michel, M. (2014). An exploratory study into trade-off effects of complexity, accuracy, and fluency on young learners' oral task repetition. *TESL Canada Journal*, *31*(8), 23-46. http://doi.org/10.18806/tesl.v31i0.1185

Schoonen, R., Snellings, P., Stevenson, M., & van Gelderen, A. (2009). Towards a blueprint of the foreign language writer: The linguistic and cognitive demands of foreign language writing. In R. Manchón (Ed.), *Writing in foreign language contexts: Learning, teaching, and research* (pp. 77-101). Bristol: Multilingual Matters.

Storch, N. (2005). Collaborative writing: Product, process, and students' reflections. *Journal of Second Language Writing*, *14*, 153-173. http://doi.org/10.1016/j.jslw.2005.05.002

Storch, N. (2007). Investigating the effectiveness of pair work on a conversational cloze task in EFL classes. *Language Teaching Research*, *11*(2), 143-159. http://doi.org/10.1177/1362168807074600

Storch, N. (2011). Collaborative writing in L2 contexts: Processes, outcomes, and future directions. *Annual Review of Applied Linguistics*, *31*, 275-288. http://doi.org/10.1017/S0267190511000079

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

Storch, N. (2016). Collaborative writing. In R. M. Manchón & P. Matsuda (Eds.), *Handbook of second and foreign language writing* (pp. 387-406). Berlin: De Gruyter Mouton.

Storch, N. (2018). Written corrective feedback from sociocultural theoretical perspectives: A research agenda. *Language Teaching*, *51*, 262-277. http://doi.org/10.1017/s0261444818000034

Storch, N., & Wigglesworth, G. (2007). Writing tasks: Comparing individual and collaborative writing. In M. P. Garcia-Mayo (Ed.), *Investigating tasks in formal language learning* (pp. 157-177). London: Multilingual Matters.

Swain, M. (2005). The output hypothesis: Theory and research. In E. Hinkel (Ed.), *Handbook on research in second language teaching and learning* (pp. 471-483). Mahwah, NJ: Lawrence Erlbaum Associates.

Swain, M., & Lapkin, S. (1998). Interaction and second language learning: Two adolescent French immersion students working together. *Modern Language Journal*, *82*, 320-337. http://doi.org/10.1111/j.1540-4781.1998.tb01209

Tavakoli, P. (2014). Storyline complexity and syntactic complexity in writing and speaking tasks. In H. Byrnes & R. M. Manchón (Eds.), *Task-based language learning: Insights from and for L2 writing* (pp. 217-236). Amsterdam: John Benjamins Publishing Company.

Teng, M. F. (2020). The effectiveness of group, pair and individual output tasks on learning phrasal verbs. *The Language Learning Journal*, *48*, 187-200. http://doi.org/10.1080/09571736.2017.1373841

Wigglesworth, G., & Storch, N. (2009). Pair versus individual writing: Effects on fluency, complexity and accuracy. *Language Testing*, *26*(3), 445-466. http://doi.org/10.1177/0265532209104670

Williams, J. (2012). The potential role(s) of writing in second language development. *Journal of Second Language Writing*, *21*(4), 321-331. http://doi.org/10.1016/j.jslw.2012.09.007

Wolfe-Quintero, K., Inagaki, S., & Kim, H. (1998). *Second language development in writing: Measures of fluency accuracy and complexity*. Honolulu, HI: University of Hawaii at Manoa.

APPENDIX

Holistic rating scale

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 organized 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 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.
2. 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.
1. This is a poorly written text. It is poorly organized and difficult to follow. Description and narration are poor or absent.