Research Article

Izaskun Villarreal* and Amaia Martínez-Sánchez

Exploring immediate and prolonged effects of collaborative writing on young learners' texts: L2 versus FL

https://doi.org/10.1515/iral-2022-0062
Received March 18, 2022; accepted December 22, 2022; published online January 19, 2023

Abstract: Context or the teaching/learning environment has only recently been recognized as a mediating variable in L2/FL written performance. These studies are multisite and have not yet targeted another feature of context: the sociolinguistic status of the target language. Likewise, scarce research exists examining the prolonged effects of collaboration. The present classroom-based study fills this void by investigating the effects of collaboration on the (a) jointly written texts; (b) subsequent individual texts; and (c) texts written in two distinct sociolinguistic status target languages of two groups of 11–12-year-old Spanish primary education students. Distributed into a control (CG) (N = 17) and an experimental group (EG) (N = 10 pairs), they wrote three descriptive texts in each language, L2 Basque and FL English: the first and third individually and the second one individually by the CG and in pairs by the EG. The texts were examined qualitatively with a rubric and quantitatively for fluency and accuracy measures. Immediate and prolonged effects of collaboration were observed on accuracy, while fluency decreased and global qualitative scores varied very little. Additionally, unlike in the CG, language-dependent differences were not attested in the EG which suggests that collaborative writing is an expedient tool to increase attention to language and limit the mediating effects of the learning context.

Keywords: collaborative writing (CW); context; English as a foreign language (EFL); quantitative & qualitative measures; second language (L2)

***Corresponding author: Izaskun Villarreal**, I-Communitas, Public University of Navarre, Campus Arrosadía, Dpto. Filología y Didáctica de la Lengua, 31006 Pamplona, Spain,
E-mail: izaskun.villarreal@unavarra.es. https://orcid.org/0000-0002-8196-5968
Amaia Martínez-Sánchez, I-Communitas, Public University of Navarre, Pamplona, Spain

1 Introduction

A growing number of school-age children across the world are afforded the opportunity to acquire some language competence in a second (L2) and/or a foreign (FL) language through the provision of different bilingual or multilingual education programmes. L2 and FL language learning contexts, however, have been considered different in terms of both the quantity and the quality of the input students receive (García Mayo and García Lecumberri 2003; Muñoz 2006), as well as in the contact with the language in and outside school. Moreover, although using productively the FL has been associated with higher FL proficiency scores, a mere increase in the quantity and quality of input has been suggested to be insufficient to develop language skills (García Mayo and Villarreal 2011). Students need to be offered opportunities for authentic discussions in and about language to thrive in language learning (Long 1996; Swain 1985). Collaborative writing (CW) or the activity by which two or more students co-construct and co-own a written text (Storch 2013) has been shown to foster such opportunities inasmuch as it encourages negotiation over language while constructing meaning, considered a source of language learning and development (Storch 2013; Swain and Watanabe 2012).

Increasingly, studies have demonstrated that CW helps to develop learners' written performance in L2 and FL contexts alike (e.g., Storch 2013 for a summary; Villarreal and Gil-Sarratea 2020 for EFL; Wigglesworth and Storch 2009 for ESL). Recently, studies have taken up to examine how learner-related variables interact with task conditions and shape the outcomes of learners' written performance, although few have involved young learners. Patterns of interaction (Azkarai and Kopinska 2020), the engagement level (García Mayo and Azkarai 2016), learners' dispositions (Kopinska and Azkarai 2020; Sato 2017), or context (Sato and Storch 2020; Storch and Sato 2019, for adults) have been recognized to mediate students' outcomes.

Following Sato and Storch (2020), context can be described as "a teaching/learning environment that governs L2 learners' needs and purposes for learning the language, including (1) the sociolinguistic status of the target language in a given community, and (2) instructional settings where the learners are situated. (p. 3)". Across-context studies have started to show that context-related variables influence students' performance including learners' interactional behaviour, output production, L1 use, uptake or spoken accuracy (Azkarai and Oliver 2016; García Mayo and Milla 2021; Llinares and Lyster 2014; Sato and Storch 2020; Storch and Sato 2019; Vold 2022). The differences have been attributed to varying proficiency levels (Azkarai and Oliver 2016), the approaches to L2 instruction (Storch and Sato 2019) or the communicative orientation of the programmes (Llinares and Lyster 2014; Vold 2022). All these studies, however, have taken place in different sites (but see

Milla 2017) and have not considered a commonplace educational setting for many school-age children: bilingual or multilingual educational settings where two or more languages with different sociolinguistic status are involved. Furthermore, none has addressed students' written outcomes.

The need to raise multilingual citizens that can face the challenges posed by the current globalized society has brought about the burgeoning of such bi or multilingual programmes. As teaching practices and learners' outcomes seem to vary depending on these languages' status and roles (Vold 2022) in the given community and instructional setting (Storch and Sato 2019), there exists a need to explore which pedagogical practices can be leveraged to improve the competence and skills acquired by these learners. To this end, the current study investigates the effects of CW in the L2 (Basque) and FL (English) of a group of Spanish 11–12-year-old primary school learners by comparing the quality of the descriptive texts produced in pairs and individually in the two languages. The results will help to inform SLA research by exploring the interplay between context (understood as in Storch and Sato 2019, p. 3) and collaboration and its effects on the developing systems of young learners, a population which is notoriously underresearched (Enever 2018; García Mayo 2017).

2 Literature review

The social constructivist theory of learning (Vygotsky 1978) establishes that cognitive and linguistic progress occur in socially embedded contexts through scaffolded interaction. L2 and FL learners have been shown to offer carefully attuned support to one another to thrive in learning (De Guerrero and Villamil 2000; Donato 1994; Swain 2000). In fact, recent studies have started to show that young learners, despite their developing metalinguistic abilities, can and do help their peers (Muñoz 2017; Tellier and Roehr-Brackin 2017; Villarreal and Munarriz-Ibarrola 2021). Their linguistic focus, however, seems to shift from aspects like spelling or word combinations to more grammatical or discursive ones, as they grow in cognitive maturity (Muñoz 2014) or proficiency (Talib and Cheung 2017).

These negotiations over language forms, known as LREs (Swain and Lapkin 1998), have resulted in better quality texts operationalized as improvements in one or various discourse analytic measures of complexity, accuracy and fluency (CAF), in global qualitative measures and/or in the functional adequacy of the written texts.

For instance, Bueno-Alastuey and Martínez de Lizarrondo Larumbe (2017) compared texts written individually, in pairs or groups regarding CAF measures in a cohort of 12–13-year-olds. They concluded that the joint texts were more accurate. Lack of improvements in complexity or fluency (words per text) measures were attributed to the text type -a descriptive text- and varying proficiency levels across the groups.

Villarreal et al. (2021) also obtained superior results after collaboration with a group of 12–13-year-old learners. The authors contrasted an individually written composition against two different CW texts (a description and a letter) which varied in the writing mode (pen and paper vs. computer-based). They obtained that the joint texts were, overall, more accurate and fluent and of higher quality, although differences were more marked for one of the tasks, the letter-writing task. The authors acknowledged that using different tasks might have affected their results.

In another study with a similar population, Villarreal and Munarriz-Ibarrola (2021) explored how grouping type (individuals, pairs and groups of three) affected the narrations written by 12–13-year-olds. Collaborative texts were of a higher quality but grouping-type was shown to mediate different constructs: pairs were more fluent and wrote more words, but small groups led to fewer lexical and grammatical errors and higher scores on global scales. The analysis of the negotiations, operationalized as language-related episodes (Swain and Lapkin 1998), also revealed that students were mainly concerned with mechanics which the authors related to maturational and cognitive developmental issues (Muñoz 2014). Similarly, they highlighted the need to investigate whether these advantages persisted in subsequent individual texts.

Finally, in the only study including a control group, Lázaro-Ibarrola (2021) investigated the impact of models on texts written by 10–11-year-old primary students. For CAF measures, the author noticed a very small effect in the CW texts and a tendency for greater accuracy and lexical diversity in the individual ones. The global scale contrasts, however, revealed greater improvements for the pairs. She encouraged the use of multiple measures and suggested that CAF measures alone might be insufficient to describe the global improvements made by students at such a young age.

All these studies share some characteristics: (1) they have been conducted with young populations; (2) they have reported linguistic benefits of CW; and (3) they explore the potential benefits of CW for the texts written in collaboration and not for its effects on posterior individual written outcomes. Investigations exploring the effects of CW on subsequent individual texts are scarce. Those examining adult populations have hinted at the effectiveness of CW not only for improving jointly written texts but also for improving posterior individual texts (e.g., Chen 2019; McDonough and De Vleeschauwer 2019; Storch 2002a, 2002b). Establishing whether CW affects L2 development and how CW shapes posterior individual language outcomes is fundamental for school children since texts written individually are the main means by which students' ability is assessed for academic credit. To the best of our knowledge, the only study exploring this on young learners is that of Bueno-Alastuey et al. (2019). These authors contrasted the initial and final individual texts – two news articles – of a group of 14–15-year-old Spanish EFL learners after CW

or peer review interventions. During the intervention, the CW group wrote a text in pairs for which they received no feedback while students in the peer assessment group wrote the texts individually and were provided with a rubric for assessing their assigned classmate's work. The two individual texts written before and after the intervention were scrutinized for CAF and holistic quality ratings and the results contrasted. They concluded that all the final texts were better than the first ones, but that the nature of the treatment mediated the results: the peer review intervention group enhanced the lexical diversity and accuracy of the texts more robustly, while the CW group wrote more fluent and complex texts but the improvement in accuracy or lexical range was more modest. This study, however, lacked jointly written texts against which to compare treatment effectiveness and to decide whether the post-test results reflected enduring gains or rather, task repetition effects (Hidalgo and Lázaro-Ibarrola 2020).

Considering that little is known about the effects of CW on young learners' jointly written texts and on subsequent individual texts as well as about how collaboration interplays with context (as in Sato and Storch 2020, p. 3) in shaping language outcomes, this study examines the texts written by a group of children in their L2, Basque, the main language of instruction in the Basque-medium immersion school students attended, and in their FL, English, the language whose use is generally limited to the English classroom. These are the research questions that guided the study:

(a) Does pair collaboration improve the quality of texts written in English and Basque in terms of quantitative (accuracy errors and words) and qualitative (adequacy, coherence, cohesion, grammar, mechanics, lexicon) parameters?
(b) Are the potential benefits retained in subsequent individual texts?
(c) Are there distinct effects on the quantitative and qualitative measures depending on the sociolinguistic status of the language they write in, such as L2 or FL?

3 Methodology

3.1 Participants and their language profiles

Participants were 49 11–12-year-old English as a foreign language (EFL) learners enrolled in a school-wide Basque total immersion programme at a semi-private school located in a mid-sized town in Northern Spain. Learners were in their sixth and last year of Primary Education.

This Northern Spain community with a population of approximately 640,000 is divided into three linguistic zones after the passing of the Basque Language Law

(1986): the Basque-speaking zone (48,938 inhabitants over-16 years of age), the mixed zone (302,428 inhabitants over-16 years of age), and the non-Basque speaking zone (182,236 inhabitants over-16 years of age). The school is located in the mixed zone in a predominantly Spanish-speaking environment in which, according to the sixth Sociolinguistic Survey (Department of Culture and Language Policy of the Basque Government 2019), around 11.3% of the population speaks Basque fluently and 12% are passive bilinguals or can speak Basque well or fairly well. These proportions increase notably as the age decreases and 27.3% are considered bilinguals in the 16–24 age range (there are no data available for younger populations). The survey revealed that the vast majority (93.8%) of citizens in this zone learnt a language other than Basque at home and among the Basque speakers, only 8.4% affirmed to be more proficient in Basque, while 71% of the Basque speakers expressed to be more fluent in their other language, namely Spanish. Regarding the use of Basque, only 0.6% of the population in this zone uses more Basque than other languages and 85.9% claims to use always a language other than Basque (i.e., Spanish or any other language they speak). Although Basque knowledge has increased in the region, Spanish is the language that is most widely used among friends, family and in formal situations. In fact, about 35% of the population in the region manifested a negative opinion towards Basque promotion, while a similar amount expressed a positive attitude towards promoting Basque when applying for a job, on TV programmes, or in education.

At the outset of the study, students were administered a sociolinguistic background questionnaire. The results were aligned with those reported in the sixth sociolinguistic survey, although a somewhat higher use of Basque was attested among the participants in the study. This was not surprising because the study was undertaken at a Basque immersion school and schools hold a fundamental role as transmitters and promotors of Basque in the community (Kasares 2013). Notwithstanding this, Basque was mostly confined to the school context and considered the mother language of very few learners. The questionnaire showed that 33 (67.3%) of the learners' first language was Spanish, while five learners considered (10.2%) Basque their mother tongue even though some of them admitted speaking in Spanish to their parents or siblings. Regarding English, this language is not part of the community and its use is mostly limited to the actual language class. The answers to the questionnaire also confirmed the foreign sociolinguistic status of English and no student expressed they used English at home or school. Almost all of them (N = 48, 97.8%), however, reported using English to an unknown extent to do playful and recreational after-school activities such as reading, listening to music, watching films or series, navigating the Internet and playing videogames. Moreover, 61.2% of students (N = 30) reported attending extra-English classes outside the school for about 1 h or 2 h per week. Yet, only three students had an English official certificate, 2 hold an A1 certificate and 1 an A2 certificate, which is in accordance with the

beginner level of the students. The questionnaire, therefore, confirmed the L2 and FL sociolinguistic status of Basque and English, respectively, and highlighted the homogeneity of the group's L2 and EFL language profiles. Results from the sociolinguistic questionnaire will not be further discussed.

The total immersion programme learners attended leads to high functional language proficiency in Basque and, as reported by the teachers, students did not experience any additional difficulties when learning curriculum contents in Basque. Regarding their English language profile, according to the English teacher, at the time of data collection, all the students had been learning English at the school for 7 years having three lessons of English per week every year which amounts to about 750 h of classroom contact time. They were reported to have a beginner level as measured by the Common European Framework of Reference for Languages (Council of Europe 2001).

Students came from two distinct intact classes with different teachers but shared the same English teacher who used identical materials and followed the same methodology for teaching them. Students followed a skill-based project approach in the English and Basque courses as well as in the rest of the courses. This approach is put into practice through the EKI Project (Ikaselkar n.d.) where basic general skills are developed in different areas of learning. In the various language arts classes, the target language is not merely the object of study, but also a vehicle to do a task, for example, learning how to be a chef or writing a magazine. Thus, students construct their own knowledge by taking part in engaging and meaningful tasks involving cross-curricular collaborative projects. So, cooperative learning and participation are encouraged. Although basic interaction among students in and about the target languages was usual, students were not used to CW practices understood as the co-authoring and co-ownership of a single text by two or more writers (Storch 2013).

At the outset of the study, learners were given an English placement test (UCLES 2019) that confirmed what the teacher said: students had a beginner English level. According to test results, 11 students had a pre-A1 level, 29 were A1 users, eight were A2 and only one was B1. A Chi-Square test determined that there were no initial differences between the groups ($\chi^2 = 6.293$; df 3; $p = 0.098$). Hence, the division into experimental and control groups was determined at random. Besides, as it seems that young learners tend to collaborate more and more effectively among similar language ability pairs (Imaz Agirre and García Mayo 2020; Villarreal and Munarriz-Ibarrola 2021), the placement results were taken into account to create the dyads in the EG. 13 parallel level pairs were formed: three pre-A1-level pairs, six A1-level pairs and four A2-level pairs. For various reasons, not all students wrote all the texts and, thus, the students that missed one of the individual texts were eliminated from the sample in the CG. However, not to compromise further the already small sample size in the experimental task, those students were considered in the EG ($N = 20$; 10 pairs). The final pool was formed by 20 participants (10 pairs) in the EG and 17 in the CG.

3.2 Materials and procedure

Six different materials were used for data collection. The tasks, the composition of descriptive texts, were chosen in collaboration with the Basque and English teachers and were designed by the second author. Descriptions were chosen because students were familiar with the type of text and topics (McDonough and Crawford 2020) in the two relevant languages and they are common in the primary education curriculum (Decreto Foral 60/2014).

The study involved eight different sessions and was carried out as part of the regular coursework over a period of five weeks. Table 1 features the procedure.

In the first session, the sociolinguistic background questionnaire was administered. In the second session, students took the placement tests (UCLES 2019) that took a maximum of 20 min. A 35-min review session on descriptions of people in English followed the test.

In the third session, all students composed the first description, the English pre-tests. To have initial measures, the texts were written individually and asking questions to the teacher or a peer was not allowed. Learners had to write their own description and were given 30 min to do it. All students finished the task in the allotted time.

In the fourth session, the students wrote the English experimental task, the second descriptive text. In this case, the CG wrote individually while the EG composed their texts in pairs. Students in pairs worked face to face and were asked to collaborate to write their joint text in response to the written prompt by generating ideas, deliberating about how to organise these ideas and deciding how best to express them (Storch 2019). The students who wrote individually had a maximum of

Table 1: Procedure for data collection.

Week	Session	Task	Grouping	Max. time
1	1	Initial questionnaire	Individual	20 min
1	2	Cambridge Placement Test & English instructions	Individual	55 min
2	3	English pre-test	Individual	30 min
2	4	English experimental task	CG: individual; EG: in pairs	45 min
3	5	English post-test & Basque instructions	Individual	55 min
4	6	Basque pre-test	Individual	30 min
4	7	Basque experimental task	CG: individual; EG: in pairs	45 min
5	8	Basque post-test	Individual	30 min

30 min, while pairs had a maximum of 45 min as previous research had reported that pairs need more time to complete tasks (Fernández Dobao 2012). However, no pair needed more than 30 min. Students were asked to describe one of their classmates.

In the fifth session, students wrote the English post-tests in which they had to describe a member of their family individually. Students were allowed 30 min and they finished the task within the time limit. For the remaining 25 min of the class, students revised descriptions of places in Basque.

In the sixth session, all students composed the first Basque description, the Basque pre-test. To have initial measures, the texts were written individually. Learners were given 30 min and wrote a description of the best holiday spot.

In the seventh session, as in the English procedure, the students wrote the Basque experimental task in which they had to describe the best place to live. The CG wrote it individually while the EG composed their texts in pairs. The students who wrote individually had a maximum of 30 min, while pairs had a maximum of 45 min. However, no pair needed more than 30 min. The same pairs from the English writing tasks were maintained in the Basque tasks.

In the eighth and final session, students wrote the Basque post-tests individually. They were asked to describe the best place for festivals or/and parties and once again, they had 30 min to do it.

Data for the study were gathered from session 3 to session 8. Written samples were not collected on consecutive days and no more than two descriptions were written in the same week. Randomizing the order of the language of the tasks was not possible as each group was formed by students from the same class, a limitation that needs to be acknowledged.

3.3 Data analysis

The data gathered in this study included a corpus of 190 texts. The CG wrote 102 individual texts, 51 in Basque and 51 in English, 17 for each task and language. The EG wrote 68 individual texts (34 in Basque and 34 in English) and 20 texts in pairs (10 in Basque and 10 in English): 17 in the pre-test, 10 in the experimental task and 17 in the post-test in each language.

To gain a more comprehensive understanding of the benefits of CW on students' written outcomes overcome the limited ecological validity of using exclusively discourse analytic measures (McDonough and García-Fuentes 2015; Villarreal and Gil-Sarratea 2020; Villarreal and Munarriz-Ibarrola 2021) as well as to align the evaluation of the texts with current pedagogical views and practices (Ball et al. 2015), the texts were analysed using both quantitative and qualitative features. Firstly, the descriptive texts were analysed in terms of quantitative measures of accuracy and

fluency (Storch 2005; Villarreal and Munarriz-Ibarrola 2021). Complexity was not analysed as the shared communicative goal of the compositions might not require any complexity growth (Pallotti 2009; Villarreal and Gil-Sarratea 2020). Fluency was measured by the total number of words produced in each text, while text accuracy was analysed by focussing on grammatical, lexical and mechanical errors (see Examples [a]–[c], respectively) as well as the total errors made. Grammatical errors included syntactic and morphological errors; lexical errors included confusion of word choice; and mechanical errors included spelling, punctuation, and capitalization. Means and the standard deviations were calculated for the analysis of the error types (see results in Table 2).

Example (a):

I live in xxxx, in a house yellow [sic]

I like play football with my friends [sic]

Example (b):

Joar has brown eyes and long hear [sic]

Example (c):

Ay live in xxxx [sic]

He likes basketball and fridays [sic]

Following Villarreal and Munarriz-Ibarrola (2021), compositions were assessed in terms of qualitative measures of adequacy, coherence, cohesion, grammar, mechanics and lexicon on a three-score band analytic rubric which was used to

Table 2: Means and standard deviations (in brackets) of accuracy errors and words in the Basque texts by task and group.

Criteria	Control group (CG)			Experimental group (EG)		
	Pre-test: ind.	Exp. task: ind.	Post-test: ind.	Pre-test: ind.	Exp. task: CW	Post-test: ind.
Grammar	3.4 (2.2)	3.9 (2.4)	4.6 (2.6)	5.1 (3.5)	4 (2.2)	4.3 (2.1)
Lexicon	1.4 (1.1)	1.6 (1.2)	2.3 (0.9)	2.1 (1.3)	2.9 (1.4)	1.6 (1)
Mechanics	7.2 (4)	6.4 (4)	7.5 (4.4)	5.4 (2.8)	5.8 (3)	3.6 (2.2)
Total errors	12 (6)	12 (5.4)	14.4 (6.5)	12.6 (5.9)	12.7 (5.4)	9.6 (3.6)
Words	70.8 (28.7)	66.6 (28.3)	56.5 (19.7)	74.1 (28.7)	52.3 (15.6)	47.1 (21)

Exp., experimental; Ind., individually; CW, in pairs.

assess both the English and Basque descriptions (see Appendix A for the rubric). Adequacy targeted the coverage of the topics and the appropriateness of the extension of the texts; coherence reviewed the clarity and comprehensibility of the compositions; cohesion assessed the organization, sequencing and development of ideas; grammatical accuracy measured the correctness of the verb tenses, subject-verb agreement, word order, pronouns, articles and prepositions; mechanics assessed the rules on spelling, punctuation and capitalization; and finally, lexicon dealt with vocabulary range (see results in Table 3).

Regarding the statistical analysis of the results, the data were introduced into an Excel spreadsheet and analysed using SPSS v26 for Windows. The normality distribution of the data varied among variables so the more conservative non-parametric tests were used for data analyses. Wilcoxon signed ranked tests were used to calculate differences between the tasks in each group as well as differences between the Basque and English results within a group. Mann-Whitney U tests were carried out to contrast the results between groups. The significance level was set at $\alpha = 0.05$.

4 Results

4.1 Quantitative and qualitative results: pre-test, experimental task and post-test comparisons

Research question one dealt with the potential differences brought about by collaboration on students' L2, Basque, and FL, English, texts. The next two sections present the inter-group comparisons for the three tasks in the L2 and FL, respectively.

Table 3: Means and standard deviations (in brackets) of qualitative measure ratings in the Basque texts by task and group.

Criteria	Control group (CG)			Experimental group (EG)		
	Pre-test: ind.	Exp. task: ind.	Post-test: ind.	Pre-test: ind.	Exp. task: CW	Post-test: ind.
Adequacy	2.2 (0.9)	1.9 (0.8)	1.7 (0.7)	1.9 (0.8)	1.6 (0.5)	1.4 (0.6)
Coherence	2.7 (0.6)	2.6 (0.8)	2.6 (0.7)	1.9 (0.8)	2.8 (0.4)	2.8 (0.4)
Cohesion	1.6 (0.9)	1.5 (0.8)	1.4 (0.6)	1.8 (0.8)	1.5 (0.7)	1.8 (0.8)
Grammar	2.3 (0.7)	2.4 (0.8)	2.1 (0.6)	2.4 (0.5)	2.2 (0.4)	2.3 (0.6)
Mechanics	2.1 (0.8)	2.1 (0.9)	1.8 (0.7)	2 (0.8)	2.1 (0.7)	1.9 (0.7)
Lexicon	2.2 (0.6)	2.2 (0.6)	1.7 (0.7)	2 (0.6)	1.7 (0.7)	1.8 (0.7)
Total	13.1 (3.3)	12.8 (3.4)	11.4 (2.6)	12 (2.8)	11.9 (2.1)	11.9 (1.8)

Exp., experimental; Ind., individually; CW, in pairs.

The Quality of The Texts Written in Basque by Task. First, Table 2 contains the results for the quantitative measures in the Basque texts. The texts did not reveal any significant difference between the groups at pre-test. The two groups made similar means of errors, although the EG produced slightly more grammatical errors (+1.7 in grammatical error means) and the CG made more mechanical errors (+1.8 in mechanical error means). Texts were also parallel in length, about 70 words.

In the experimental task, the texts written individually by the CG and in pairs by the EG revealed significant differences regarding lexical errors ($Z = -2.188$; $p = 0.029$): the CG produced a lower mean of errors (1.6) than the EG (2.9). None of the other comparisons turned out to be significant, and once again, both groups exhibited similar accuracy means. As for fluency, the CG wrote (non-significantly) longer texts than the EG (+14.3 words), approximately 66 words.

At post-test, the individual texts showed that the EG wrote shorter (–9.4 words) but more accurate texts than the CG. Statistically significant differences favouring the EG were attested: the EG made significantly fewer total accuracy errors (–4.8), as well as mechanical errors in their subsequent texts (–3.9) ($Z = -2.262$; $p = 0.024$ and $Z = -2.759$; $p = 0.006$, respectively). Grammatical and lexical accuracy measures had similar means.

Regarding qualitative results, Table 3 features the mean scores for the quality measures of the Basque texts.

The pre-test texts received very similar ratings, but the CG wrote texts that were more coherent than the EG and the difference turned out to be statistically significant (+0.8, $Z = -2.997$; $p = 0.003$).

For the experimental task, the CG exhibited a statistically significantly richer lexical repertoire and obtained a mean that was 0.5 higher than that of the pairs ($Z = -2.052$; $p = 0.040$). The rest of the ratings were very similar as was the mean of the total ratings.

The results for the post-test were consistent too and no notable differences were attested between the groups, 1.4 was the lowest score for the two groups (in cohesion and adequacy), while both groups got the highest ratings for coherence (2.6 the CG and 2.8 the EG).

All in all, the analyses of the Basque texts revealed some differences between groups and tasks. While the accuracy of the texts was comparable at pre-test, the EG made more lexical accuracy errors in the experimental task, a difference that was statistically significant. The post-test results for the accuracy measures, on the other hand, revealed that the EG wrote texts that were statistically significantly more accurate in mechanical aspects as well as in the means of total errors. The comparison of the qualitative ratings, on the other hand, did not unveil advantages for collaborative writing in the experimental task or post-test. In fact, significant

differences for the CG regarding the measures of coherence at pre-test and lexical errors in the experimental task were obtained.

The Quality of The Texts Written in English by Task. Table 4 shows the means of accuracy error and words for the texts written in English. No significant differences emerged between the groups when all students wrote their first descriptions individually, at pre-test. Students in the EG wrote somewhat longer texts (+5 words) and obtained slightly higher error means, +1.2 errors in mechanics and +1.8 in total errors.

In the experimental task, the pair texts were found to be significantly more grammatically accurate ($Z = -2.528$; $p = 0.011$) than those written individually by the CG. On average, they contained 3.3 fewer errors. The other contrasts did not yield significant differences, and it was the CG who wrote longer texts (+12.1 words) and produced slightly fewer mechanical errors (–1.2).

In their subsequent individual texts, in the post-test, the EG made fewer grammatical and total errors (–1.2, –2.1, respectively). In contrast, the CG made slightly fewer mechanical errors (–0.4) and wrote longer texts (+2.5). None of the contrasts, however, turned out to be significant.

As was the case for the Basque qualitative ratings, the English qualitative ratings were very similar across groups and tasks (Table 5) and no contrast revealed significant variations between the groups. In the pre-test, the rating for mechanics and the total rating were modestly higher for the CG (+0.5 and +0.8), while the EG obtained somewhat higher results in grammar (+0.4).

In the experimental task, both groups did better than in the pre-test and no notable differences emerged between them. Coherence was the highest for both groups (2.8 CG and 2.7 EG), while grammar (1.6) was the lowest for the CG.

The ratings for the individual post-tests did not vary much, but the EG obtained higher scores in 5 out of the 7 criteria: coherence, cohesion, mechanics, lexical and

Table 4: Means and standard deviations (in brackets) of accuracy errors and words in the English texts by task and group.

Criteria	Control group (CG)			Experimental group (EG)		
	Pre-test: ind.	Exp. task: ind.	Post-test: ind.	Pre-test: ind.	Exp. task: CW	Post-test: ind.
Grammar	4.8 (3)	6.9 (3.4)	5.4 (3.8)	4.9 (3.3)	3.6 (2.2)	4.2 (1.9)
Lexicon	1.7 (1.7)	2.5 (1.6)	2.4 (1.7)	2.3 (1.7)	2.4 (1.7)	1.6 (1)
Mechanics	4.1 (3.3)	3.1 (2.1)	3.5 (2.2)	5.3 (2.8)	4.3 (3)	3.9 (2.6)
Total errors	10.6 (4.4)	12.4 (3.9)	11.8 (5.5)	12.4 (5.9)	10.3 (5)	9.7 (3.8)
Words	67.53 (31.3)	65.5 (32.3)	50.4 (25.8)	72.7 (16.5)	53.4 (15.9)	47.9 (15.2)

Exp., experimental; Ind., individually; CW, in pairs.

Table 5: Means and standard deviations (in brackets) of qualitative measure ratings in the English texts by task and group.

Criteria	Control group (CG)			Experimental group (EG)		
	Pre-test: ind.	Exp. Task: iInd.	Post-test: ind.	Pre-test: ind.	Exp. Task: CW	Post-test: ind.
Adequacy	2.1 (0.9)	2.1 (0.8)	1.9 (0.8)	2.1 (0.9)	1.7 (0.8)	1.7 (0.6)
Coherence	2.8 (0.6)	2.8 (0.7)	2.5 (0.8)	2.6 (0.8)	2.7 (0.7)	2.8 (0.6)
Cohesion	1.8 (0.8)	1.3 (0.5)	1.5 (0.8)	1.7 (0.8)	1.7 (0.8)	1.8 (0.9)
Grammar	1.8 (0.7)	1.6 (0.7)	2 (0.5)	2.2 (0.9)	2 (0.7)	1.9 (0.8)
Mechanics	2.4 (0.7)	2 (0.7)	1.9 (0.7)	1.9 (0.7)	2.1 (0.7)	2.2 (0.7)
Lexicon	2.2 (0.8)	2.1 (0.8)	1.8 (0.7)	1.8 (0.7)	1.7 (0.8)	2.1 (0.6)
Total	13.1 (2.7)	11.8 (2.8)	11.7 (3.1)	12.3 (2.6)	11.9 (3.3)	12.4 (2.3)

Exp., experimental; Ind., individually; CW, in pairs.

the total score, but the magnitude of the difference was small (0.7 in the total score and 0.3 in the rest). The CG only obtained slightly higher marks in two out of the seven criteria, cohesion and grammar, with a 0.2 and 0.4 mean score gain respectively.

Overall, regarding the English texts, the quantitative results evinced a significant effect for collaboration only for grammatical accuracy: the EG pair texts were grammatically more accurate. No significant differences were attested between the groups at pre or post-test. The EG, however, made fewer errors than the CG in the post-test except for mechanical errors while the EG made more errors across categories in the pre-test. Qualitative measures were strikingly similar and no differences between groups were obtained. Additionally, the results showed that the pairs were non-significantly less fluent than the individuals and that their texts were also shorter when they wrote individually in the post-test.

4.2 Enduring effects of CW: pre-test versus post-test results by group

The second research question explored retention in subsequent students' texts.

The Control Group. For Basque quantitative measures, repeating the task did not seem to promote accuracy and fluency (see Table 2). All the error means were higher in the post-test, although only the grammatical error comparison turned out to be significant (3.4 vs. 4.6, $Z = -2.159$; $p = 0.031$). The CG produced significantly more grammatical errors at post-test than at pre-test. Fluency contrasts supported the trend and the first Basque text was significantly longer than the post-test texts (–24.3, $Z = -2.226$; $p = 0.026$).

Likewise, qualitative measures (see Table 3) exhibited a similar pattern and the pre-test texts were rated the highest. The comparisons revealed that the third and final texts were rated significantly lower than the experimental texts, written individually in the CG (12.8 vs. 11, $Z = -3.119$; $p = 0.002$), and the pre-tests (13.1 vs. 11.4, $Z = -2.781$; $p = 0.005$) and were considered less adequate than the pre-test (2.2 vs. 1.7, $Z = -2.530$; $p = 0.011$).

Regarding the results for the English texts, a similar picture emerges. The CG was non-significantly less accurate and fluent (see Table 4) in the individual post-tests, except for mechanics (4.1 vs. 3.5). The fluency contrasts yielded significant differences as students wrote –17.1 fewer words in the post-test ($Z = -2.249$; $p = 0.025$).

Task repetition benefits were not reflected in qualitative measures either. The CG obtained lower ratings in all the measures but in grammar, which improved marginally (1.8 vs. 2). Their post-test ratings were overall statistically significantly poorer (13.1 vs. 11.7, $Z = -2.684$; $p = 0.007$) as were the lexical range ($Z = -2.333$; $p = 0.020$) and mechanics ($Z = -2.126$; $p = 0.033$).

The Experimental Group. The pre-post-test contrasts regarding results for the Basque quantitative measures (see Table 2) demonstrated that the texts written by the EG after the collaborative practice were invariably more accurate, although they did not reach significance. In contrast, they wrote significantly fewer words than they had done in the pre-test ($Z = -2.226$; $p = 0.026$).

The ratings for the Basque qualitative measures (see Table 3) also declined for the EG, except for cohesion and total ratings which did not vary. In fact, adequacy and coherence contrasts turned out to be significant ($Z = -2.333$; $p = 0.020$; $Z = -3.035$; $p = 0.002$, respectively) which demonstrated that EG students' texts were less adequate but more coherent at post-test than at pre-test.

The English pre-post-test quantitative contrasts showed a non-significant drop in accuracy errors (see Table 4). The subsequent individual texts contained fewer grammatical (–0.7), lexical (–0.7), mechanical (–1.4), and total errors (–2.7). The fluency contrast yielded significant results as students wrote 24.8 fewer words ($Z = -3$; $p = 0.003$).

The English qualitative rating comparisons unfolded a more complex scenario (see Table 5). The differences between the pre-test and the post-test were very subtle and non-significant. While students obtained higher rating means for coherence (+0.2), cohesion (+0.1), mechanics (+0.3), lexicon (+0.3), total ratings (0.1), they presented a minor decrease in adequacy (–0.4) and grammar (–0.3) ratings.

In short, the CG did not improve the quality of the texts written in Basque and English. What is more, the third texts were the shortest and poorest, although only the fluency measure revealed statistically significant differences. The final texts of the EG, on the other hand, revealed non-significant accuracy improvements but were significantly shorter. Regarding qualitative ratings, these ratings did not show

any clear improvement, and EG student ratings lowered in Basque and increased marginally in English.

4.3 Quantitative and qualitative text differences: L2 versus FL

Research question three targeted the effect of sociolinguistic status on the development of quantitative and qualitative measures. In what follows, the results for the CG are presented first followed by the results for the EG.

The Control Group. When looking at the results obtained for the two languages regarding quantitative measures, fluency scores showed remarkable similar scores in the two languages. No length restrictions were set for the descriptions, but the first texts in the L2 and the FL were about 70 words, while the third and final ones were much shorter, about 56 words.

Some accuracy means seemed to be affected by language, though. While students made similar amounts of lexical errors, they produced notably more grammar errors in English (4.8, 6.9 and 5.4) than in Basque (3.4, 3.9 and 4.6), although only the difference in the experimental task was significant ($Z = -2.542$; $p = 0.011$). Mechanical errors disclosed significant differences regardless of the task. Basque texts contained more mechanical errors (7.2, 6.4 and 7.5) than the English ones (1.7, 2.5 and 2.4; pre-test: $Z = -2.803$; $p = 0.005$; experimental task: $Z = -3.001$; $p = 0.003$; post-test $Z = -3.222$; $p = 0.001$).

The qualitative ratings did not vary widely between the languages, although significant differences emerged for grammar at pre-test ($Z = -2.309$; $p = 0.021$) and in the experimental task ($Z = -2.697$; $p = 0.007$). The Basque texts by the CG were rated +0.5 higher at pre-test and +0.8 higher during the second task. These differences disappeared at post-test and they obtained similar means in the ratings, 2 in English and 2.1 in Basque.

The Experimental Group. The fluency count was remarkably similar across tasks and languages. Students wrote a similar amount of words in the L2 and the FL: about 74, 52 and 47 in Basque and 73, 53 and 50 in English.

Accuracy means contrasts also did not disclose differences in the two languages for the EG, although in the experimental task they reduced their mechanical error means slightly more in English (4.3) than in Basque (5.8) as did the total accuracy means (10.3 and 12.7 respectively).

Similarly, no variation was observed for the qualitative aspects. The ratings were strikingly similar across languages and tasks and the magnitude of the variation tended to be smaller than 0.3. The broadest difference was on the pre-test cohesion, in which the English texts were rated +0.7 higher.

Therefore, the contrast between the L2 and FL measures disclosed group differences. While no noticeable variation was attested in the EG, the CG produced more mechanical errors in the Basque L2 while significant grammatical accuracy and holistic differences emerged in the first two tasks, but not in the post-test.

5 Discussion

This study explored the effects of CW on the text quality of jointly written texts and subsequent individual texts and how CW interplays with the sociolinguistic status of the target language in which the texts were written. The study was deemed necessary, as little evidence exists regarding what benefits writing collaboratively might bring about to the developing L2 competence of primary learners. Furthermore, how the local context of learning and, in particular, the sociolinguistic status of the languages involved, interplays with collaboration was also examined because, albeit underresearched, it is commonplace for many children enrolled in multilingual programmes.

The first research question addressed differences in accuracy and fluency measures and global quality scales between the collaborating group, the EG, and the CG. Fluency seemed to be unaffected by collaboration and both groups followed a similar pattern in the three tasks and the two languages. Students wrote less in each task. This contrasts with investigations undertaken with similar cohorts (e.g., Bueno-Alastuey and Martínez de Lizarrondo Larumbe 2017; Lázaro-Ibarrola 2021; Villarreal and Munarriz-Ibarrola 2021) which have detected non-significant increases in text length. The similarity of the pattern detected between the control and experimental groups suggests that an effect related to the nature of the text type, descriptions, might be at play. Students might have become more efficient and fulfilled the task more competently (Storch 2005), in fact, even though pairs have been said to take longer (Fernández Dobao 2012), all students finished their tasks before the allotted time. Possible motivational effects cannot be ruled out either. Insights from these young learners' motivational dispositions could help us to determine their influence on task completion (e.g., Dörnyei and Kormos 2000; Kopinska and Azkarai 2020; Villarreal and Lázaro-Ibarrola 2022), a learner-related factor to be considered for future investigation.

Likewise, global qualitative scores did not unveil differences between the groups, and no benefits of CW were observed in the experimental task. It might be the case that the analytic rubric used, originally conceived for assessing narrations (Villarreal and Munarriz-Ibarrola 2021) might have lacked sufficient descriptive power to grasp the differences between the various descriptions. A functional adequacy scale, as in Hidalgo and Lázaro-Ibarrola (2020) and Lázaro-Ibarrola (2021),

which aims at a more global evaluation of writing might have helped detect variations among the texts.

Our study, however, did reveal a statistical advantage in grammatical accuracy in favour of the EG in their FL texts. Our results regarding accuracy mirror previous findings concerning young EFL learners (e.g., Bueno-Alastuey and Martínez de Lizarrondo Larumbe 2017; Villarreal and Munarriz-Ibarrola 2021; Villarreal et al. 2021). This might be due to the language-focused discussions our students engaged in while writing together which resulted in higher accuracy rates as students drew their attention to form when they negotiated meaning (Swain and Watanabe 2012; Villarreal and Gil-Sarratea 2020; Villarreal and Munarriz-Ibarrola 2021). The lack of improvements in the English mechanical accuracy means might be attributed, following Villarreal et al. (2021), to the learners' developing language awareness abilities and the instability of the English language of these beginner learners (Muñoz 2014) or to decision making difficulties experienced in collaboration (Villarreal and Munarriz-Ibarrola 2021). No accuracy differences were observed in the Basque L2 texts in the experimental task, however. Students' high functional proficiency in the L2 or the meaning-oriented instruction (García Mayo and Milla 2021; Vold 2022) might induce students to prioritize meaning over form and, as they struggle less, they create fewer opportunities for negotiations over language when constructing meaning and therefore, correct their texts less. Unfortunately, recording students' dialogues was not possible due to privacy issues, a shortcoming that is acknowledged here.

The second research question delved into the effects of CW on subsequent individual texts. The findings echoed those from the experimental task: no fluency or qualitative differences were detected for any of the groups, but modest accuracy effects emerged. This is in line with Chen (2019) and Bueno-Alastuey et al. (2019) who obtained positive persisting effects in accuracy rates after collaboration, after extended opportunities for language negotiation, which resulted in higher grammatical accuracy. Our CG obtained its weakest accuracy means in both languages at post-test, except for the FL mechanical scores, whereas the EG evidenced a slight drop in error rates which suggests that the language-focused negotiations undertaken during the collaborative written process might have fostered increased attention to accuracy also in the subsequent individual written task. Moreover, the EG exhibited a somewhat more accurate performance in the two languages. The positive trend attested in the post-test FL texts coupled with the statistical decrease of mechanical and total accuracy errors obtained in the L2 post-test texts lend strong support to the positive contribution of collaborative practices to accuracy enhancement.

The third research question explored the interplay between CW and the local context, namely the sociolinguistic status of the languages involved. The findings

obtained highlight the role of CW to counterbalance (Lyster and Mori 2006; García Mayo and Milla 2021) students' priorities when writing in Basque, their L2, and English, their FL. No statistical differences between the two languages were attested for the EG, although such differences emerged in the CG. The CG made significantly more mechanical errors in Basque across tasks and more grammatical errors in English, which seems to reflect students' higher Basque proficiency as well as their attention focus. Immersion programs aim at developing students' L2 proficiency in the service of academic development (Ballinger et al. 2017) and therefore, prioritize meaning over form. In accordance with Llinares and Lyster (2014) and Sato and Storch (2020), the findings mirror language-dependent school practices, but the lack of differences attested in the EG suggests that CW might serve to counterbalance (García Mayo and Milla 2021; Lyster and Mori 2006) the lack of attention to form in immersion programmes. The CG does not focus on mechanical issues in Basque as much as they do in English, because these errors do not hinder communication in Basque and functional proficiency is key in such programmes. Most probably, such corrections are more explicit in the FL (Llinares and Lyster 2014; Vold 2022) triggered by students' lowest English proficiency but also by the opaque English spelling system. Integrating CW practices seems therefore to be a promising tool to increase form-focused (Long 1991) instruction authentically as it seems to direct students' attention to linguistic forms.

6 Conclusion

Our study shows that collaborative writing improves text accuracy and that this benefit seems to persist in a posterior individual attempt which can be suggestive of language development. In addition, CW might help to counterbalance students' attention to form and meaning (García Mayo and Milla 2021; Lyster and Mori 2006). The findings of this study provide evidence for teachers to encourage CW practices in language classes by suggesting that CW can help students to improve their individual writing ability, to pay increased attention to language, and to progress towards their ultimate goal of language development (Chen 2019).

Some pedagogical suggestions can be drawn based on our findings. Writing in pairs seems to increase attention to form and help to improve the accuracy of texts even in posterior individual attempts. Therefore, collaborative writing practices with YL should be encouraged to draw genuine attention to form that will foster language development (Chen 2019; Nassaji 2016) especially in L2 settings (García Mayo and Milla 2021; Vold 2022). CW seems to serve as an expedient tool to counterbalance students' meaning-oriented performance (Lyster and Mori 2006) and attract attention to form, a balance that seems necessary to thrive in language learning.

This study has some shortcomings that should be acknowledged. The lack of recordings during the experimental tasks only allows for speculation regarding the focus of students' dialogues and their linguistic consequences. Further work would be needed to describe the nature, frequency and outcome of such dialogues (Villarreal and Gil-Sarratea 2020; Villarreal and Munarriz-Ibarrola 2021), to map out the cognitive processes students follow in their L2 and FL (see, for instance, Torres and Cung 2019 for L2 and Heritage Language Learners or López-Flamarique et al. 2022 for the L1), to investigate the potential crosslinguistic influence and to determine what characteristics of collaboration are beneficial. Furthermore, a delayed post-test would allow to rule out possible task repetition effects (Hidalgo and Lázaro-Ibarrola 2020; Manchón 2014) and explore enduring CW effects (Chen 2019). In addition, the fact that the study examined the effects of a one-off treatment might have masked the genuine effects that ongoing CW pedagogical activities might bright about in language classrooms. Future longitudinal studies would be helpful to investigate whether ongoing engagement in CW could lead to additional linguistic advantages. Finally, although the study carries high ecological validity since the educational programme described is a relatively frequent multilingual programme, the social status of the languages investigated and students' exposure to them could be considered confounding variables. Future studies, therefore, should try to compare current results with the result from a setting in which one of the two variables can be controlled for. Similarly, future studies should control for the proficiency of students in the languages involved by, perhaps, giving students initial placement tests that would serve as an additional argument to establish the status of those languages. Notwithstanding, the findings of this study underscore the value of CW for L2 and FL learning and encourage teachers to include collaborative practices in their language classes. Whether and how to introduce and train students to make full use of their linguistic repertoire through translanguaging practices that maximize the effectiveness of CW merits further investigation (e.g., Chen 2019; Kim and McDonough 2011; Seals et al. 2020). This study has opened up a relatively uncharted research agenda and investigating the extent to which CW hastens the language learning process of multilingual school-age children seems worth pursuing.

Acknowledgments: We would like to thank the school, the teachers and the children for their willingness and generosity to participate in this project.

Appendix A

Analytic rubric for assessing the compositions based on Villarreal and Munarriz-Ibarrola (2021).

		Marks		
		3	2	1
Task	Adequacy	– All the points in the instructions are mentioned. – The length of the text is appropriate.	– Some points in the instructions are not mentioned. – The text could be longer: Ideas could be more developed.	– Notable omission of the content points and/or considerable irrelevance of some of them. – The text is too short: Ideas are not fully developed.
	Coherence	– The text is clear. – The text is easy to understand.	– The text is clear. – The text is easy to understand, although some incoherent points can confuse the reader.	– The text is not clear. – The text is difficult to understand.
Language	Cohesion	– Ideas are well organised. – The use of paragraphs is adequate. – Cohesive devices linking sentences and paragraphs are used.	– Ideas are organised. – The paragraphs could be better distributed. – Some cohesive devices linking sentences and paragraphs are used.	– There is a lack of organisation. – Paragraphs are too long or there are not well distributed, or there are no paragraphs at all. – There are no linking devices.
	Grammatical accuracy	– Very few, irrelevant, or no grammar errors at all. – Good command of grammar: Correct use of the verb to be and the third person.	– Some acceptable grammar errors. – Fair command of English grammar: use of the verb to be and third person (with some mistakes).	– Serious and numerous grammar mistakes. – Does not use correctly the verb to be and the third person.
	Mechanics	– Most words are written correctly. – There are only some occasional mistakes.	– Some spelling mistakes (3–6). – Some mistakes in basic vocabulary.	– There are many spelling mistakes. – Invents words.
	Lexical range	– Rich and varied vocabulary.	– Basic vocabulary, enough to convey the message.	– Limited range of vocabulary: some words are in Basque-Spanish.

References

Azkarai, Agurtzane & Marta Kopinska. 2020. Young EFL learners and collaborative writing: A study on patterns of interaction, engagement in LREs, and task motivation. *System* 94. 102338.

Azkarai, Agurtzane & Rhonda Oliver. 2016. Negative feedback on task repetition: ESL vs. EFL child settings. *Language Learning Journal* 47(3). 1–12.

Ball, Phil, Keith Kelly & John Clegg. 2015. *Putting CLIL into practice*. Oxford: Oxford University Press.

Ballinger, Susan, Roy Lyster, Andrea Sterzuk & Fred Genesee. 2017. Context-appropriate crosslinguistic pedagogy. *Journal of Immersion and Content-Based Language Education* 5(1). 30–57.

Bueno-Alastuey, Camino, Maríá & Rodero Albaiceta Sheila. 2019. The effects of using collaborative writing vs. peer review treatments on subsequent individual writings. *Huarte de San Juan. Filología y Didáctica de la Lengua* 193. 32–61.

Bueno-Alastuey, María Camino & Patricia Martínez de Lizarrondo Larumbe. 2017. Collaborative writing in the EFL Secondary Education classroom comparing triad, pair and individual work. *Huarte de San Juan. Filología y Didáctica de La Lengua* 17. 254–275.

Chen, Wenting. 2019. An exploratory study on the role of L2 collaborative writing on learners' subsequent individually composed texts. *The Asia-Pacific Education Researcher* 28. 563–573.

Decreto Foral 60/2014, de 16 de julio, por el que se establece el currículo de las enseñanzas de Educación primaria en la Comunidad Foral de Navarra [Regional Decree 60/2014, of July 16, which establishes the curriculum of primary education in the Chatered Community of Navarra]. Boletín Oficial de Navarra. Pamplona, 5 September 2014, number 174, pp. 1–181.

Council of Europe. 2001. *Common european framework of reference for languages: Learning, teaching, assessment*. Cambridge: Cambridge University Press. Available at: http://www.coe.int/t/dg4/linguistic/source/Framework_EN.pdf.

De Guerrero, María & Olga S. Villamil. 2000. Activating the ZPD: Mutual Scaffolding in L2 Peer Revision. *The Modern Language Journal* 84. 521–568.

Department of Culture and Language Policy of the Basque Government. 2019. *Sixth sociolinguistic survey 2016*. Vitoria-Gasteiz: Eusko Jaurlaritzen Argitalpen Zerbitzu Nagusia.

Donato, Richard. 1994. Collective scaffolding in second language learning. In James P. Lantolf & Appel Gabriela (eds.), *Vygotskian approaches to second language research*, 33–56. Norwood: Ablex.

Dörnyei, Zoltan & Judit Kormos. 2000. The role of individual and social variables in oral task performance. *Language Teaching Research* 4(3). 275–300.

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

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

García Mayo, María Pilar (ed.). 2017. *Learning foreign languages in primary schools. Research insights*. Bristol: Multilingual Matters.

García Mayo, María Pilar & Agurtzane Azkarai. 2016. EFL task-based interaction: Does task modality impact on language-related episodes? In Masatoshi Sato & Susan Ballinger (eds.), *Peer interaction and second language learning. Pedagogical potential and research agenda*, 241–266. Amsterdam: John Benjamins Publishing Company.

García Mayo, María Pilar & Izaskun Villarreal. 2011. The development of suppletive and affixal tense and agreement morphemes in the L3 English of Basque–Spanish bilinguals. *Second Language Research* 27(1). 129–149.

García Mayo, María Pilar & María Luisa García Lecumberri. 2003. *Age and the acquisition of English as a foreign language*. Clevedon: Multilingual Matters Ltd.

García Mayo, María Pilar & Ruth Milla. 2021. Corrective feedback in second versus foreign language contexts. In Hossein Nassaji & Eva Kartchava (eds.), *The Cambridge handbook of corrective feedback in second language learning and teaching*, 473–494. Cambridge: Cambridge University Press.

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

Imaz Agirre, Ainara & María Pilar García Mayo. 2020. The impact of agency in pair formation on the degree of participation in young learners' collaborative dialogue. In Craig Lambert & Rhonda Oliver (eds.), *Using tasks in second language teaching: Practice in diverse contexts*, 306–323. Bristol: Multilingual Matters.

Ikaselkar. n.d. Eki proiektua [The Eki project]. http://www.ikaselkar.eus/en/eki/eki-information/?atala=ezaugarri-bereizgarriak-en (accessed 12 March 2021).

Kasares, Paula. 2013. Euskararen belaunez belauneko jarraipena eta hizkuntza sozializazioa Nafarroako familia euskaldunetan (1970–2012) [Generational transmission and linguistic socialisation of the Basque language in Basque-speaking families in Navarre (1970–2012)]. *Fontes Linguae Vasconum* 116. 209–233.

Kim, YouJing & Kim McDonough. 2011. Using pretask modelling to encourage collaborative learning opportunities. *Language Teaching Research* 15(2). 183–199.

Kopinska, Marta & Agurtzane Azkarai. 2020. Exploring young EFL learners' motivation: Individual vs. pair work on dictogloss tasks. *Studies in Second Language Learning and Teaching* 10(3). 607–630.

Lázaro-Ibarrola, Amparo. 2021. Models in collaborative and individual writing among EFL children: Noticing, incorporations and draft quality. *International Review of Applied Linguistics in Language Teaching*.

Llinares, Ana & Roy Lyster. 2014. The influence of context on patterns of corrective feedback and learner uptake: A comparison of CLIL and immersion classrooms. *Language Learning Journal* 42(2). 181–194.

Long, Michael H. 1991. Focus on form: A design feature in language teaching methodologyKees de Bot, Ralph B. Ginsberg, & Claire Kramsch. In de Bot Kees, Ralph B. Ginsberg & Claire Kramsch (eds.), *Foreign language research in cross-cultural perspective*, 39–52. Amsterdam: John Benjamins.

Long, Michael H. 1996. The role of the linguistic environment in second language acquisition. In William C. Ritchie & Tej K. Bhatia (eds.), *Handbook of second language acquisition*, 413–468. San Diego: Academic Press.

López-Flamarique, Maite, Mónica Aznárez-Mauleón & Isabel García del Real. 2022. ¿Cómo planifica y revisa el alumnado de primaria? Un estudio a partir de una tarea de escritura colaborativa. *Didáctica. Lengua y literatura* 35.

Lyster, Roy & Hirohide Mori. 2006. Interactional feedback and instructional counterbalance. *Studies in Second Language Acquisition* 28(2). 269–300.

Manchón, Rosa María. 2014. The distinctive nature of task repetition in writing. Implications for theory, research, and pedagogy. *Estudios de Lingüística Inglesa Aplicada ELIA* 14. 13–41.

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

McDonough, Kim & Jindarat De Vleeschauwer. 2019. Comparing the effect of collaborative and individual prewriting on EFL learners' writing development. *Journal of Second Language Writing* 44(April). 1–8.

McDonough, Kim & William J. Crawford. 2020. Identifying effective writing tasks for use in EFL write-to-learn language contexts. *Language Learning Journal* 48(4). 469–480.

Milla, Ruth. 2017. *Corrective feedback episodes in CLIL and EFL classrooms: Teachers' and learners' beliefs and classroom behaviour*. Vitoria, Spain: University of the Basque Country (UPV/EHU) Unpublished doctoral thesis.

Muñoz, Carmen (ed.). 2006. *Age and the rate of foreign language learning*. Clevedon: Multilingual Matters.

Muñoz, Carmen. 2014. Exploring young learners' foreign language learning awareness. *Language Awareness* 23(1–2). 24–40.

Muñoz, Carmen. 2017. The development of language awareness at the transition from primary to secondary school. In María Pilar García Mayo (ed.), *Learning foreign languages in primary school: Research insights*, 49–68. Bristol: Multilingual Matters.

Nassaji, Hossein. 2016. Anniversary article: Interactional feedback in second language teaching and learning: A synthesis and analysis of current research. *Language Teaching Research* 20(4). 535–562.

Pallotti, Gabrielle. 2009. CAF: Defining, refining and differentiating constructs. *Applied Linguistics* 30(4). 590–601.

Sato, Masatoshi. 2017. Interaction mindsets, interactional behaviors, and L2 development: An affective-social-cognitive model. *Language Learning* 67(2). 249–283.

Sato, Masatoshi & Neomy Storch. 2020. Context matters: Learner beliefs and interactional behaviors in an EFL vs. ESL context. *Language Teaching for Young Learners* 26(5). 919–942.

Seals, Corinne A., Jonathan Newton, Madeline Ash & Bao Trang Thi Nguyen. 2020. Translanguaging and task based language teaching: Crossovers and challenges. *Educational Linguistics* 45(August). 275–292.

Storch, Neomy. 2002a. Patterns of interaction in ESL pair work. *Language Learning* 52(1). 119–158.

Storch, Neomy. 2002b. Relationships formed in dyadic interaction and opportunity for learning. *International Journal of Educational Research* 37(3–4). 305–322.

Storch, Neomy. 2005. Collaborative writing: Product, process, and students' reflections. *Journal of Second Language Writing* 14(3). 153–173.

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

Storch, Neomy. 2019. Collaborative writing as peer feedback. In Ken Hyland & Fiona Hyland (eds.), *Feedback in second language writing (contexts and issues)*, 143–162. Cambridge: Cambridge University Press.

Storch, Neomy & Masatoshi Sato. 2019. Comparing the same task in ESL vs. EFL learning contexts: An activity theory perspective. *International Journal of Applied Linguistics (United Kingdom)* 30(1). 1–20.

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

Swain, Merrill. 2000. The output hypothesis and beyond: Mediating acquisition through collaborative dialogue. In James P.2 Lantolf (ed.), *Sociocultural theory and second language learning*, 97–114. Oxford: Oxford University Press.

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

Swain, Merrill & Yuko Watanabe. 2012. *The encyclopedia of applied linguistics*. *Social, dynamic and complexity theory approaches to second language acquisition*. Wiley Blackwell. https://doi.org/10.1002/9781405198431.wbeal0664.

Talib, Tarmizi & Yin Ling Cheung. 2017. Collaborative writing in classroom instruction: A synthesis of recent research. *The English Teacher* 46(2). 43–57.

Tellier, Angela & Karen Roehr-Brackin. 2017. Raising children's metalinguistic awareness to enhance classroom second language learning. In María Pilar García Mayo (ed.), *Learning foreign languages in primary school: Research insights*, 22–48. Bristol: Multilingual Matters.

Torres, Julio & Bianca Cung. 2019. A comparison of advanced heritage language learners' peer interaction across modes and pair types. *The Modern Language Journal* 103(4). 815–830.

UCLES, Cambridge English Placement test. 2019. *Test your English*. Cambridge English assessment. Available at: https://www.cambridgeenglish.org/test-your-english/for-schools/.

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

Villarreal, Izaskun & Miren Munarriz-Ibarrola. 2021. "Together we do better": The effect of pair and group work on young EFL learners' written texts and attitudes. In María Pilar García Mayo (ed.), *Working collaboratively in second/foreign language learning*, 89–115. Berlin, Boston: De Gruyter Mouton.

Villarreal, Izaskun, María Camino Bueno-Alastuey & Raquel Sáez-León. 2021. Computer-based collaborative writing with young learners: Effects on text quality. In María Pilar García Mayo (ed.), *Working collaboratively in second/foreign language learning*, 89–115. Berlin, Boston: De Gruyter Mouton.

Villarreal, Izaskun & Amparo Lázaro-Ibarrola. 2022. Models in collaborative writing among CLIL learners of English in primary school: Linguistic outcomes and motivation matters. *System* 110. 1–21.

Vold, Eva Thue. 2022. Learner spoken output and teacher response in second versus foreign language classrooms. *Language Teaching Research*. *Online first* https://doi.org/10.1177/13621688211068610.

Vygotsky, Levy. 1978. *Mind in society: The development of higher psychological processes*. Cambridge, MA: Harvard University Press.

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