Studies in English Language Teaching
ISSN 2372-9740 (Print) ISSN 2329-311X (Online)
Vol. 11, No. 1, 2023
www.scholink.org/ojs/index.php/selt

*Original Paper*

# Exploring Peer Collaborative Writing in EFL Secondary Mixed-age Classrooms

Dr. Tomas Kos[1] & Friedrich, Zoe Lara[1]

[1] English and American Studies, Humboldt University of Berlin, Berlin, Germany

Received: February 9, 2023 Accepted: February 25, 2023 Online Published: February 27, 2023

http://doi.org/10.22158/selt.v11n1p55 URL: http://dx.doi.org/10.22158/selt.v11n1p55

***Abstract***

*Based on the beliefs (assertions) that students benefit socially and cognitively from interacting with peers of different ages, mixed-age (M-A) classrooms have attracted much attention from educators, education policymakers, and researchers. Nevertheless, research into the role and impact of age grouping in second (L2) and foreign language (FL) learning is rather scarce. Moreover, a growing body of research has shown that by linking oral and written modes, collaborative writing (CW) aids language learning. However, the majority of studies have investigated adult and young adult students, and very few were conducted with secondary and high school students. Based on our knowledge, no study has explored CW in M-A secondary school classrooms. With these research gaps in mind, and drawing predominantly on documentary analysis of student written work, the present study examined CW of 28 students of English as a foreign language (EFL) aged 12 to 15 old learners who jointly wrote a comic. The findings point to limitations of M-A peer interaction suggesting that there is a threshold for the younger learners in terms of benefiting from CW with their older partners. Similarly, the findings suggest that the extent to which OS benefit is rather limited.*

***Keywords***

*collaborative writing, mixed-age peer interaction, secondary EFL classrooms, CAF*

## 1. Introduction

Mixed-age (M-A) classrooms (also referred to as multi-grade, cross-age, or composite) comprise about one-third of all classrooms worldwide (Saqlain, 2015). They are composed of students of different ages and two, three, or even four different grades and vast differences in terms of abilities are very common (Smit & Engeli, 2015). Classrooms composed of students of different ages (commonly referred to as multi-grade) have been established long ago in rural areas around the world because of a lack of students or financial difficulties (Saqlain, 2015). Recent years have seen an increase in schools that set

up M-A classrooms based on the beliefs of the school community in the positive pedagogical and social outcomes of this approach (Ronksley-Pavia, Barton, & Pendergast, 2019). This has been the case in public and private schools in countries such as Australia, Canada, Germany, the Netherlands, Sweden, France, and Switzerland (Thoren & Brunner, 2019; Saqlain, 2015). For example, M-A classrooms have been implemented in alternative education such as the *Jena plan*, grounded in the pedagogy of Petersen, or in Montessori schools which advocate learning in heterogeneous groups and underline the importance of creating opportunities for children to interact and work with peers of different ages (Pape, 2016). In Germany, M-A classrooms seem to be also popular in *Gesamtschulen* which are public schools combining three school types (primary, secondary, and high school) into a comprehensive school (Pape, 2016). The term M-A is used in this study to refer to different ages, grades, and relative proficiency, while the term same-age (S-A) refers to the same grade, similar age, and relative proficiency. Likewise, unless specified, to avoid repetition, the term *older student(s)* (OS) is used when referring to older students who were simultaneously more proficient learners within pairs while the term *younger student(s)* (YS) refers to younger learners who in this study were also the less proficient learners.

The proponents of M-A classrooms have asserted that interacting with peers of different ages aids their intellectual development (Hoffman, 2002; Little, 2001; Pape, 2016). The assumption is that the YS benefit from being tutored or mentored by their older peers, while the OS students benefit from teaching the younger ones (Little, 2001; Wagener, 2014). Nevertheless, despite the increase of schools that have implemented FL teaching in M-A classrooms, the research to test these assumptions is still limited (Kos, 2019a, b; Frank, 2014; Heinzmann et al., 2015; Thurn 2011). This is particularly the case for peer interactions among young adolescent students (Kos, 2019a, b). Given that M-A pair/group work is one of the pillars on which M-A classrooms stand, as a great deal of work is meant to be done by students in interactions with their younger and older peers, language teachers in M-A classrooms need to know to what extent such grouping is beneficial to learning. Moreover, given that many assignments in secondary school years involve writing, it is important to explore to what extent collaborative writing (CW) among M-A pairs aids to language learning of each student within the pair. Although research on CW has indicated learning benefits for young, young adolescent and adult learners (Calzada & Garcia Mayo, 2020a, b, 2021; Storch, 2011, 2013, 2016; Villarreal & Gil-Sarratea, 2019), studies on CW among peers of different ages are missing. Based on my knowledge, no study has investigated CW among M-A peer interaction in the secondary school context. Moreover, as most studies on peer interaction have been conducted outside students' familiar environment while using tasks that are not commonly performed during regular lessons (e.g., *dictogloss*), the current study explored CW on a common classroom task to get a more accurate picture of the potential of CW in a regular FL classroom. It investigated the difference between students' performance during CW and subsequent individual writing (IW) performance, which were both assessed for academic credit. This is particularly important for M-A classrooms as it aims to shed some light on to what extent M-A peer interaction aids the subsequent IW process and the written use of linguistic features targeted by the task.

This study involved 28 young adolescent learners aged between 12 and 15 and focused on a comic writing task. This article reports findings that were part of a larger study that investigated peer interactions in M-A classrooms.

## 2. Literature Review

### *2.1 Related Research on M-A and Mixed-proficiency Peer Interaction*

M-A peer interaction is generally grounded in the sociocultural theory of mind which views social interaction between an *expert* and *novice* as a crucial space for the child's development because it provides the child with structures that he/she internalizes in later stages as cognitive capacities (Vygotsky, 1978). In other words, learning and development occur in learners' interaction with people in his/her environment such as in cooperation with his/her peers and the more capable peer's help is crucial for the less capable peer to perform a task which he/she is not able to perform individually. Some researchers have claimed that grouping children across ages reduces antisocial behavior (Hoffman, 2002) and promotes cooperative behavior, sharing, help, and self-directed learning (Wagener, 2014). Not all research, however, supports these claims. Although some studies have reported cognitive, social, and emotional benefits of M-A classrooms (Katz, Evangelou, & Hartman, 1990; Wagener, 2014), some have found no differences in the academic performance of M-A vs. S-A learners (Gutiérrez & Slavin, 1992; Quail & Smyth, 2014). What is more, some studies have shown negative effects on social and cognitive development (Huff & Raggl, 2015; Lindström & Lindahl, 2011). Findings from mainstream education have explored peer tutoring among M-A (referred to as cross-age) and S-A peers and their effects on learning gains are not clear-cut either. Although some studies are supportive of the M-A grouping (Topping & Bryce, 2004), some seem to contradict the assertion that heterogenous grouping is necessary for children's cognitive development and academic performance (Robinson, Schofield & Steers-Wentzell, 2005). Some suggest that interactions among peers of equal or similar abilities may be more conducive to learning because they are reciprocal (Duran & Monereo, 2005).

Research related to L2 teaching and learning in M-A classrooms in secondary school classrooms is still scarce. Kos (2019a, b) has explored how secondary school peers of differing ages relate to each other when interacting on classroom tasks and how their established relationships impact the nature and the extent of their assistance and language learning. For example, when interacting on classroom tasks, some students assisted one another in ways similar to teacher scaffolding, while some in ways that resemble what Donato (1994) called *collective scaffolding* (Kos, 2019a). In another study, Kos (2019b) has shown that despite age/proficiency differences, M-A pairs may form patterns of interaction, which are conducive to learning, namely *expert/novice*/*collaborative* pattern (Storch, 2002). These studies have shown that, although some students may benefit from their peers' help, the lowest-grade students of low proficiency, in particular, may not. Moreover, it is the relationship among peers, as well as the perceptions of the partner's proficiency and the learners' goals that greatly impact how peers interact with one another.

*2.2 Research on Collaborative Writing*

CW refers to the writing process in which two or more writers interact and jointly produce a text (Storch, 2016). Storch (2011) has defined CW as the co-construction or co-Kosship of a text by two or more writers. The notion of co-Kosship (joint ownership) refers to "a sense of shared responsibility and ownership of the text, and thus to the substantial participation of all co-Koss in all phases of text production" (Storch, 2021, p. 14). This shared responsibility for a written text implies that students have a common goal to achieve, are interdependent, help each other, and must work closely together while pooling their resources and ideas and communicating thoughts, intentions, and feelings to each other (Storch, 2021). For example, they "negotiate and agree on what ideas to include, how to organize those ideas, and how best to express those ideas" (Storch, 2021, p. 14). They may as well be settling down disagreements, thus improving their communication and social skills.

Another rationale for CW is that it combines some of the advantages of writing and speaking. While writing is characterized by a slow pace and visible results, speaking involves the availability of an "audience" and immediate feedback. Thus, CW can be beneficial for FL teaching because it is slower, allows for visible output, and can, therefore, provide for deeper engagement with language (Williams, 2012). Perhaps as a consequence, CW offers opportunities to use language structures that may not be commonly used in oral communication (Williams, 2012). And because writers can see their text as they write, and writing is less spontaneous and immediate than speaking, students may feel less anxious than with purely oral communication (Tavakoli, 2014). Studies with YL have shown positive effects of CW tasks on grammar development (Calzada & Garcia Mayo, 2021, 2020a), motivation and patterns of interaction (Azkarai & Kopinska, 2020) or on learners' attitudes toward CW tasks such as the dictogloss (Calzada & Garcá Mayo, 2020b). Research has underlined the benefits of feedback, correction, and text reformulation in CW (Coyle & Roca de Larios, 2014) and indicated positive effects of task repetition on the accuracy of a written text, and occurrence and resolution of Language Related Episodes (LREs) (Hidalgo & Lázaro-Ibarrola, 2020), during which learners talk about the language they produce, question their language use, or correct themselves or others (Swain & Lapkin, 1998). For example, focusing on ten pairs aged 12 in an EFL classroom, Hidalgo and Lázaro-Ibarrola (2020) examined two written texts in response to the same picture prompt three times over three weeks. In their study on the effects of task repetition on the complexity, accuracy, and fluency (CAF) of the drafts. The researchers found that even though task repetition led to more accurate texts, complexity and fluency remained unchanged. However, a more holistic analysis going beyond CAF has shown that learners produced better compositions in terms of content, structure, and task fulfillment. In addition, learners were able to generate and resolve a high number of form-focused and meaning-focused LREs.

Despite the benefits of CW, teachers seem to be reluctant to use CW tasks; particularly in the secondary school context due to assessment issues (McDonough, de Vleeschauwer, & Crawford, 2018). They may prefer using CW for brainstorming ideas before the writing activity itself, or for obtaining feedback from the teacher or peers on the drafted or completed text (Storch & Wigglesworth, 2007). Nevertheless,

research on CW in the secondary school context is scarce. In a classroom-based study, Villarreal and Gil-Sarratea (2019) explored the CW of 32 Basque–Spanish bilinguals (16-17 years old) learning EFL in their first year of non-compulsory secondary education. Implementing a control (IW) and experimental group (CW), the study found that CW promoted shorter, less fluent but more accurate and slightly more lexically and grammatically complex texts than IW. Moreover, CW afforded texts of better quality in terms of content, structure, and organization of ideas as assessed by holistic measures. This is a positive finding indicating that CW can be beneficial for students regardless of proficiency. The study attempts to answer the following questions:

RQ1) To what extent do texts written collaboratively among secondary school students organized into mixed-age pairs differ from subsequently individually written texts in terms of complexity, accuracy, and fluency?

RQ2) To what extent do the same texts differ in terms of content, structure, and level of detail?

## 3. Method

### *3.1 Context and Participants*

This study took place in three EFL classrooms at an alternative school secondary school in Saxony, Germany. It was conducted over one unit of work which lasted two and a half months. The data presented in this article comes from one class attended by 28 students of 7$^{th}$, 8$^{th}$ and 9$^{th}$ grade who were 12 to 15 years old and who formed 14 M-A pairs (see Table 1 below). Before the unit of work, students were told that they were going to complete a proportion of assigned tasks included in their individual study plans with a partner. They were allowed to select a partner as long as she/he was of a different grade. They were told that together they were responsible for the completion of this plan. The completed study plans were to be submitted at the end of the unit and were graded. A follow-up classroom achievement test related to the content of the unit of work was conducted one week after the end of the unit. This article draws on the data gained from students' pieces of CW including the comic and an IW classroom achievement test.

One of the limitations of the study was that learners' language proficiency could not be assessed independently of school-based assessment. Participants' "relative proficiency" could only be assessed by two classroom achievement tests which were taken throughout the first term and by other summative classroom assessment practices which aimed to assess learners' speaking skills and grammatical and vocabulary knowledge. In addition, formative assessment practices in the form of observation of learners' performance during lessons and taking notes were considered. However, the assessment practices were specific to the grade, and therefore the description is relative to the particular grade, and not an estimate relative to overall proficiency (Kos, 2019a). In other words, a true comparison of learners' language abilities across grades is not possible. Students were assured that at every stage, their names would remain confidential and the data would be kept securely and would be used for academic purposes only.

**Table 1. Participant Characteristics**

| Pair | Name | Age | Grade | RP | Sex |
|---|---|---|---|---|---|
| 1 | Lucy/Lea | 13/12 | 8/7 | H/H | F/F |
| 2 | Jan/Anton | 15/14 | 9/8 | H/H | M/M |
| 3 | Mirek/Edie | 15/14 | 9/8 | M/M | M/M |
| 4 | Ira/Ronja | 14/13 | 8/7 | H/M | F/F |
| 5 | Jens/ Fabian | 15/14 | 9/8 | H/H | M/M |
| 6 | Georg/Jim | 14/13 | 8/7 | H/H | M/M |
| 7 | Lotta/Livia | 14/13 | 8/7 | M/M | F/F |
| 8 | Alena/Elena | 14/13 | 8/7 | H/H | F/F |
| 9 | Romy/ Lilia | 14/13 | 8/7 | M/L | F/F |
| 10 | Luna/Ruby | 14/13 | 8/7 | M/M | F/F |
| 11 | Yael/Sophie | 14/13 | 8/7 | M/M | F/F |
| 12 | John/Viktor | 15/13 | 9/7 | H/M | M/M |
| 13 | Liv/Jasmin | 15/14 | 9/8 | H/M | F/F |
| 14 | Mana/Ada | 15/14 | 9/8 | H/H | F/F |

RP: relative proficiency; H: High, M: Middle, L: Low (relative to year group as assessed by the first term assessment practices

*3.2 Task*

Toward the end of unit of work, students carried out the comic writing task in which they were asked to jointly read the comic and write it either as A) a dialog or B) a story. The task was embedded in the lessons together with examples of other related tasks and exercises as well as examples of the targeted language (See appendix A for an illustration of the task and B for lesson procedure). The comic task is a convergent task that is, a task "in which all speakers are working to a jointly agreed outcome" (Adas, 2003, p.123). The task combined reading, speaking, and writing. Research suggests that using writing and speaking tasks, rather than speaking tasks alone, would increase the amount of engagement with a language form while learners' attention is also directed to meaning (Alegria de la Colina & Garc á Mayo, 2007). It has to be noted, however, the task lacked its sole focus on meaning as it contained a grammar exercise in the pre-task phase to raise students' awareness of the targeted linguistic form before engaging in the task.

One week later, students took an individual classroom achievement test, which, in addition to other tasks, contained the comic task which was identical to the one which the students completed jointly. Nevertheless, it has to be noted that the research-based aim of the achievement test was not to make accurate predictions about learners' improvement of writing skills, because the design of the study did

not include a pre-test. However, it can shed some light on to what extent the CW in the M-A constellation could have played in students' IW.

*3.3 Analytical Procedure*

RQ1) To what extent do texts written collaboratively among secondary school students organized into mixed-age pairs differ from subsequently individually written texts in terms of complexity, accuracy, and fluency?

Both CW and IW texts were first examined for CAF. Complexity was determined by establishing the ratio of dependent clauses to all clauses and in terms of lexical diversity, which was assessed by the number of different words in a text divided by the total number of words. It needs to be mentioned that a T-unit, the shortest unit of a text that contains one independent clause along with its dependent clause/s and can be segmented without 'leaving any sentence fragments as residue' (Hunt, 1970: 189) was not included as the above-mentioned measures seemed to have been sufficient for this purpose. Accuracy was determined by establishing the proportion of error-free clauses to the total number of clauses. Also, the number of errors per total number of words was considered because it accounts for the exact distribution of errors concerning words (Storch, 2005). No distinction was made between syntactical, morphological, and lexical (confusion of word choice) or mechanical (spelling and punctuation) (Fernández Dobao, 2012). Fluency was measured in terms of the total number of words and the number of words per sentence. A paired-sample t-test was run for both groups to compare the CW and IW texts and to determine the statistical significance by tracking the difference in the means. To estimate differences across different age groups, a tally was made for each measure for each student.

**Example 1. Collaborative Writing**

(Lucy & Lea, grade 8 & 7, 144 words)

*Once upon a time there was a gang. On Saturday the gang met for mural painting. Jaden had to leave the gang, because he had a date. The gang was angry of Jaden. Jaden went to the cafe and there was Chloe. Later called Sandy Jaden and told him that the gang needed him. But Jaden say's that he can not came. Because of his answer the gang was really angry of him. So the gang destroited the mural, and went home. Later Jaden left the cafe and on his way home he saw the mural and had an idea. Jaden fix the mural and Sandy saw it. On the competition the gang won the 2nd Prize. After the competition Sandy met Jaden and Chloe in the cafe. Sandy told Jaden that in the next month is an competition. Chloe was angry about it.*

**x = error free clause x = error free dependent clause x= error**

**Table 2. Analysis of CAF Measures**

| All clauses | Dependant clauses | Error-free clauses |
|---|---|---|
| 25 | 4 | 16 |
| Number Sentences | Number of different words | Number of errors |
| 15 | 61 | 9 |

**Example 2. Individual writing**

**(Lea, grade 7, 151 words)**

Once upon a time there was a "mural gang". On Saturday they met. They've started the work, but Jaden forgot that he has a date, so he ran to the cafe but he was too late, because Chloe already waited for him. Half an hour later, Sandy phoned Jaden and told him that they needed him. He just answered that he was sorry, but he hasn't got time right now. After the call the gang was really mad at Jaden so they made the mural ugly. After Jaden's date, on his way home he saw the mural and had an idea. He worked along until the mural got finished. On the mural competition, they won the 2nd prize. After the competition,Sandy met Jaden and Chloe in the cafe. She told Jaden about her idea to do a mural together. Jaden was really happy about her idea, but Chloe was angry.

**x = error free clause x = error free dependent clause x= error**

**Table 3. Analysis of CAF Measures**

| All clauses | Dependant clauses | Error-free clauses |
|---|---|---|
| 26 | 4 | 26 |
| Number Sentences | Number of different words | Number of errors |
| 12 | 83 | 0 |

RQ2) To what extent do the same texts differ in terms of content, structure, and level of detail?

Texts were analyzed qualitatively using holistic measures of content, structure, and attention to detail with four score bands (based on Shehadeh, 2011, p. 4: very good; 3: good/average; 2: fair/poor; and 1: very poor). Content refers to the targeted relevance to the topic and quality of the text in terms of what was required, necessary, or expected. Structure relates to the parts (or lack of parts) included in the text and the clarity of ideas. Attention to detail refers to the ability to attend to specific features in writing such as precision or/and explicitness in expressing ideas using appropriate language. For example, the text above (Figure 1) was assessed as 4 on content and structure and 3 on level of detail. Similar to a) a paired-sample t-test was run for both groups and tallies were made for each measure and individual student.

Both researchers reviewed independently all pieces of writing (CW and IW texts) and reached an overall consensus in 92% of instances concerning RQ1 and 86% in relation to RQ2. Most disagreements were related to holistic measures. Later, differences were discussed and an agreement was reached.

## 4. Result

RQ1) To what extent do texts written collaboratively among secondary school students organized into mixed-age pairs differ from subsequently individually written texts in terms of complexity, accuracy, and fluency?

Table 4 below shows the overall results for CW and IW texts across CAF and content, structure, and level of detail measures.

**Table 4. Complexity, Accuracy and Fluency/Overall Results**

| | Collaborative writing | | | Individual writing | | |
|---|---|---|---|---|---|---|
| | R | M | SD | R | M | SD |
| DC/C | 0.08-0.39 | 0,20 | 0.94 | 0-0.38 | 0.23 | 0.91 |
| DW/WT | 0.15-0.65 | 0.51 | 0.93 | 0.34-0.6 | 0.52 | 0.6 |
| EFC/CT | 0.12-0.96 | 0.55 | 0.25 | 0.20-1 | 0.64 | 0.20 |
| E/WT | 0.01-0.19 | 0.09 | 0.05 | 0-0.57 | 0.09 | 0,10 |
| WT | 107-218 | 144.3 | 34.2 | 35-245 | 134.3 | 43.3. |
| W/S | 7.1-18.2 | 12.3 | 3.1 | 6.4-15.4 | 11.3 | 2.2 |

R: Range; M: Mean; SD: standard deviation; DC: dependent clause; C: clause; DW: different words; WT: words total; EFC: error free clause; CT: total number of clauses; E: error; WS: words per sentence

As shown in Table 4, with regards to complexity, IW texts had a higher dependent clause to all clauses ratio (M=0.23, SD 0.91) than CW texts (CW) (M=0.20, 0.94). This difference was statistically insignificant (t=-1.373, M=0.91). IW texts were also higher on lexical complexity (M=0.52, SE 0.6) than CW texts (M=0.51, SD= 0.93). This difference was not significant either (t=-1.323; p=0.451). On measures of accuracy, the IW texts scored higher. IW texts had a significantly higher ratio of error-free clauses to the total number of clauses (M=0.64, SD=0.25) than CW texts (M=0.55, SD=0.25) (t=-1.8181; p=0.040). Also, students in the IW condition produced fewer errors in relation to the total number of words (M=0.0896, SE= 0,103) in comparison to CW (M= 0.0921, SE= 0.054), but the difference was not statistically significant (t=-0.129; p=0.449). When looking at fluency, CW texts contained a higher number of words (M=144.3, SD=34.2) than IW texts (M=134.3, SD=43.3) and had a higher number of words per sentence ratio (M=12.3, SD=3.1) than IW texts (M=11.3, SD=2.2). None of these differences were statistically significant. To sum up, IW texts scored higher on complexity and accuracy whereas CW texts were more fluent than IW texts. Table 3 below shows results related to CAF measures for each pair and individual student.

When looking at individual students (Table 5 below), the findings reveal that on measures of complexity, nine out of 14 OS and five out of 14 YS showed higher gains on IW. There was, however, a minor difference across both measures of complexity for OS. Students had, on average, higher gains in

accuracy than complexity as 10 out of 14 OS produced more error-free clauses (EFC) and fewer errors concerning the total number of words (E/TW) when compared with CW. Similarly, 11 out of 14 YS students showed higher gains on error-free clauses (EFC) and seven out of 14 committed fewer errors concerning the total number of words (E/TW). In terms of fluency, nine out of 14 OS wrote a higher number of words, and seven out of 14 had a higher word/sentence ratio on IW than CW. In contrast, only three out of 14 YS wrote more words on IW than during CW. Finally, only half of the YS produced a higher word/sentence ratio on IW. Overall, the findings show that students, regardless of age, improved the most in accuracy when compared with CW. In addition, except for error-free clauses (EFC) and the number of words per sentence (W/S), the OS seemed to have scored higher on all other measures than their younger partners.

**Table 5. Complexity, Accuracy, and Fluency & Content, Structure, and Level of Detail Across Pairs/Individual Students**

| | | | Complexity | | | | Accuracy | | Fluency | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Name | A | DC/ CW | DC/C /IW | DW/W T/CW | DW/W T/IW | EFC/C T/CW | EFC/C T/IW | E/W/ CW | E/WT /IW | WT/ CW | WT/ IW | W/S/ CW | W/S IW |
| **1** | Lucy | 13 | 0.16 | 0.20 | 0.42 | 0.53 | 0.64 | 0.70 | 0.06 | 0.05 | 144 | 165 | 9.5 | 11.0 |
| | Lea | 12 | 0.16 | 0.15 | 0.42 | 0.55 | 0.64 | 1.00 | 0.06 | 0.00 | 144 | 151 | 9.5 | 12.6 |
| **2** | Jan | 15 | 0.08 | 0.31 | 0.56 | 0.51 | 0.25 | 0.58 | 0.19 | 0.08 | 136 | 159 | 10.5 | 14.5 |
| | Anton | 14 | 0.08 | 0.20 | 0.56 | 0.56 | 0.25 | 0.50 | 0.19 | 0.08 | 136 | 125 | 10.5 | 12.5 |
| **3** | Mirek | 15 | 0.14 | 0.21 | 0.51 | 0.57 | 0.41 | 0.63 | 0.13 | 0.08 | 166 | 97 | 14.7 | 12.1 |
| | Edie | 14 | 0.14 | 0.38 | 0.51 | 0.56 | 0.41 | 0.63 | 0.13 | 0.07 | 166 | 41 | 14.7 | 13.7 |
| **4** | Ira | 14 | 0.17 | 0.27 | 0.50 | 0.6 | 0.96 | 0.77 | 0.01 | 0.05 | 137 | 111 | 12.5 | 9.3 |
| | Ronja | 13 | 0.17 | 0.15 | 0.50 | 0.51 | 0.96 | 0.20 | 0.01 | 0.16 | 137 | 113 | 12.5 | 10.3 |
| **5** | Jens | 15 | 0.29 | 0.31 | 0.65 | 0.51 | 0.21 | 0.58 | 0.16 | 0.08 | 108 | 159 | 18.0 | 14.5 |
| | Fabian | 14 | 0.29 | 0.35 | 0.65 | 0.50 | 0.21 | 0.65 | 0.16 | 0.09 | 108 | 151 | 18.0 | 13.7 |
| **6** | George | 14 | 0.10 | 0.21 | 0.55 | 0.60 | 0.60 | 0.96 | 0.08 | 0.01 | 107 | 119 | 7.1 | 8.5 |
| | Jim | 13 | 0.10 | 0.21 | 0.55 | 0.56 | 0.60 | 0.47 | 0.08 | 0.09 | 107 | 96 | 7.1 | 10.7 |
| **7** | Lotta | 14 | 0.39 | 0.38 | 0.54 | 0.58 | 0.57 | 0.79 | 0.09 | 0.03 | 162 | 144 | 13.3 | 13.1 |
| | Livia | 13 | 0.39 | 0.34 | 0.54 | 0.54 | 0.57 | 0.62 | 0.09 | 0.07 | 162 | 160 | 13.3 | 11.9 |
| **8** | Alena | 14 | 0.29 | 0.19 | 0.50 | 0.53 | 0.58 | 0.81 | 0.11 | 0.04 | 138 | 129 | 12.5 | 9.2 |
| | Elena | 13 | 0.29 | 0.23 | 0.50 | 0.50 | 0.58 | 0.65 | 0.11 | 0.07 | 138 | 145 | 12.5 | 12.1 |
| **9** | Romy | 14 | 0.12 | 0.16 | 0.49 | 0.51 | 0.12 | 0.48 | 0.16 | 0.12 | 140 | 120 | 18.2 | 8.6 |
| **10** | Luna | 14 | 0.30 | 0.18 | 0.50 | 0.48 | 0.74 | 0.79 | 0.06 | 0.14 | 139 | 141 | 11.6 | 11.4 |
| | Ruby | 13 | 0.30 | 0.15 | 0.50 | 0.51 | 0.74 | 0.20 | 0.06 | 0.16 | 139 | 113 | 11.6 | 10.3 |
| **11** | Yael | 14 | 0.14 | 0.22 | 0.61 | 0.43 | 0.67 | 0.74 | 0.08 | 0.03 | 110 | 145 | 8.5 | 8.6 |

| | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Sophie | 13 | 0.14 | 0.20 | 0.61 | 0.46 | 0.67 | 0.76 | 0.08 | 0.05 | 110 | 130 | 8.5 | 8.7 |
| **12** | John | 15 | 0.22 | 0.33 | 0.15 | 0.51 | 0.28 | 0.67 | 0.11 | 0.06 | 107 | 143 | 11.9 | 13 |
| | Viktor | 13 | 0.22 | 0.08 | 0.57 | 0.60 | 0.28 | 0.31 | 0.11 | 0.14 | 107 | 72 | 11.9 | 12 |
| **13** | Liv | 15 | 0.14 | 0.12 | 0.55 | 0.54 | 0.74 | 0.88 | 0.03 | 0.02 | 208 | 185 | 12.2 | 15.4 |
| | Jasmin | 14 | 0.14 | 0.23 | 0.55 | 0.46 | 0.74 | 0.62 | 0.03 | 0.08 | 208 | 185 | 12.2 | 12.3 |
| **14** | Mana | 15 | 0.31 | 0.31 | 0.44 | 0.44 | 0.87 | 0.87 | 0.02 | 0.03 | 218 | 245 | 9.9 | 12.9 |
| | Ada | 14 | 0.31 | 0.25 | 0.44 | 0.44 | 0.87 | 0.69 | 0.02 | 0.06 | 218 | 180 | 9.9 | 10.0 |

A: age; S: sentence, DC: dependent clause; DW: different words; WT: words total; EFC: error free clause C: clause; CT: total number of clauses; E: error; WS: words per sentence

RQ2) To what extent do the same texts differ in terms of content, structure, and level of detail?

With regards to content, structure and level of detail, Table 6 below shows that students had significantly higher scores on the CW texts in terms of content (M=3.2, SD=0.96) than on IW texts (M=2.5, SD=1.0) (t=3.8, p<.001). CW texts were higher on text structure (M=2.92, SD=0.81) than IW texts (M=2.9, SE=0.83), but the difference was not statistically significant (t=0.21, p=.416). Finally, CW texts were significantly higher concerning the level of detail, (M=2.79; SD=0.79) than IW texts (M=2.4, SD=1.1) (t=2.4, p<.011). To sum up, on average, holistic measures indicate that students' performance on IW declined when compared with CW.

**Table 6. Content, Structure, and Level of Detail/Overall Results**

| | Collaborative writing | | | Individual writing | | |
|---|---|---|---|---|---|---|
| Content | 1-4 | 3.2 | 0.6 | 1-4 | 2.5 | 1 |
| Structure | 2-4 | 2.92 | 0.81 | 1-4 | 2.90 | 0.83 |
| Level of detail | 2-4 | 2.79 | 0.79 | 1-4 | 2.43 | 1.14 |

R: Range; M: Mean; SD: standard deviation

When looking at individual learners' performance (Table 7 below), only three students' texts improved concerning content. The texts of nine students remained the same and the texts of 15 students scored lower. Similar results can be seen with respect to the text structure as only six students' texts improved, and the level of detail on which only four students scored higher when compared with CW. When looking across ages, only two OS and five YSs' texts showed some level of improvement concerning the content when compared to CW while 10 out of 14 YSs' texts scored lower. In relation to text structure, only four OS and two YSs' texts improved, the texts of 14 (seven OS and YS) students scored the same and four OS and six YS scored lower than in the CW condition. In relation to the level of detail, only one OS and 3 YS scored higher. The texts of 10 OS and 3 YS remained the same while lower gains were seen in the case of three OS and seven YS. Overall, when looking at individual

students, their performance concerning holistic measures seemed to have declined when contrasted with CW. Also, OS scored higher than their younger partners on IW than CW.

**Table 7. Content, Structure, and Level of Detail Across Pairs/Individual Students**

| | | | Content | | Structure | | Level of detail | |
|---|---|---|---|---|---|---|---|---|
| | Name | A | C/CW | CTF/IW | ST/CW | ST/IW | LD/CW | LD/IW |
| **1** | Lucy | 13 | 4 | 3 | 4 | 3 | 3 | 3 |
| | Lea | 12 | 4 | 4 | 4 | 3 | 3 | 3 |
| **2** | Jan | 15 | 3 | 3 | 3 | 3 | 2 | 2 |
| | Anton | 14 | 3 | 2 | 3 | 3 | 2 | 3 |
| **3** | Mirek | 15 | 3 | 2 | 2 | 2 | 3 | 2 |
| | Edie | 14 | 3 | 1 | 2 | 2 | 3 | 1 |
| **4** | Ira | 14 | 4 | 3 | 4 | 3 | 2 | 1 |
| | Ronja | 13 | 4 | 2 | 4 | 2 | 2 | 1 |
| **5** | Jens | 15 | 2 | 3 | 2 | 3 | 2 | 2 |
| | Fabian | 14 | 2 | 2 | 2 | 3 | 2 | 3 |
| **6** | George | 14 | 3 | 4 | 2 | 4 | 2 | 3 |
| | Jim | 13 | 3 | 2 | 2 | 2 | 2 | 1 |
| **7** | Lotta | 14 | 4 | 4 | 4 | 4 | 4 | 4 |
| | Livia | 13 | 4 | 3 | 4 | 3 | 4 | 4 |
| **8** | Alena | 14 | 4 | 2 | 3 | 3 | 3 | 3 |
| | Elena | 13 | 4 | 3 | 3 | 3 | 3 | 2 |
| **9** | Romy | 14 | 2 | 1 | 2 | 3 | 2 | 1 |
| **10** | Luna | 14 | 2 | 1 | 2 | 1 | 2 | 1 |
| | Ruby | 13 | 4 | 2 | 3 | 3 | 3 | 3 |
| **11** | Yael | 14 | 4 | 2 | 3 | 2 | 3 | 1 |
| | Sophie | 13 | 3 | 3 | 3 | 3 | 3 | 3 |
| **12** | John | 15 | 3 | 2 | 3 | 3 | 3 | 2 |
| | Viktor | 13 | 1 | 2 | 2 | 3 | 2 | 2 |
| **13** | Liv | 15 | 1 | 1 | 2 | 1 | 2 | 1 |
| | Jasmin | 14 | 4 | 4 | 3 | 4 | 4 | 4 |
| **14** | Mana | 15 | 4 | 2 | 3 | 4 | 4 | 4 |
| | Ada | 14 | 4 | 4 | 4 | 4 | 4 | 4 |

A: age; CW-collaborative writing IW: individual writing; C: Content; ST: Structure; LD: Level of detail

## 5. Discussion

The current study aimed to explore to what extent CW among M-A pairs aid the subsequent IW and the use of grammatical features elicited by the task of individual students. To do so, it compared the difference between students' CW performance and subsequent IW writing performance, which were both assessed for academic credit. Involving 28 young adolescent learners aged between 12 and 15 and focusing on a comic writing task during regular classroom EFL lessons, this study aimed to contribute to the research body on peer interaction conducted in naturalistic classroom settings by exploring a context of M-A classrooms in which the research is still much needed.

Students' written texts were first examined employing CAF measures (RQ1) as well as more holistic measures including task content, structure, and level of detail (RQ2). When looking across CAF measures, the findings indicate that on average, students' texts gained in terms of complexity and accuracy while their performance declined in fluency. This supports Hidalgo and Lázaro-Ibarrola's (2020) findings that repeating the task may lead to more accurate texts. However, in contrast to their study, in which complexity and fluency remained unchanged with task repetition, students in the current study produced more complex but less fluent texts on IW. Nevertheless, none of these findings was statistically significant. Moreover, caution is needed when comparing these studies as the research designs differed. The findings of the current study differ from Villarreal and Gil-Sarratea's (2019) findings (note that the research designs also vary) in which CW promoted shorter and less fluent texts than IW. One possible explanation for why students' CW texts in this study were more fluent than IW texts is that students did not experience time pressure during CW which they could have during the individual achievement test. It has to be, however, noted that the time allocated for the achievement test was sufficient (90 minutes) and that the CW texts were graded as well. Another possible explanation is that during CW students were more concerned with expressing the meaning while paying less attention to syntactical and morphological aspects of language. Similar to Villarreal and Gil-Sarratea (2019), IW texts were more accurate and more complex than CW. This can be interpreted twofold. It can be argued that this is a positive finding as it suggests that CW may have promoted students' focus on language form during the subsequent IW. Alternatively, students may have attended more to the language form during the individual achievement test than during a CW activity because they simply perceived the importance of "getting it right" on the test rather than writing "more". In other words, students' perceptions of the writing activity may have influenced their purpose and their performance on IW.

When looking at age-related differences, the findings indicate that except for two measures (EFC and W/S), the OS scored higher on all other measures than their younger partners. This is not a surprising finding given the differences in the time of instruction received over the years. Undoubtedly, given that no pre-test was conducted, no accurate estimates regarding students' learning gains linked to CW can be made. Nevertheless, some tendencies can be observed. While OS tended to score on all measures to a similar extent, YS improved most on accuracy (EFC) but dropped most on fluency as only three out of 14 YS wrote more words than during CW and only half had a higher word/sentence ratio. One

possible explanation is that due to the relative difficulty of the individual achievement test for the YS, they did not have the linguistic resources to grapple with all aspects of writing and seemed to have attended the most to the accurate use of language. Similar to the above, their perception of a test as being a tool to assess accurate language use could have played a role.

Rather surprising findings come from the holistic analysis including content, structure, and level of detail (RQ2). CW texts amounted significantly higher in content and level of detail than IW texts. CW texts were also higher on structure than IW texts. These findings are in line with those of Villarreal and Gil-Sarratea (2019) and Hidalgo and Lázaro-Ibarrola (2020) which suggest that students write texts of "better quality" in terms of content, structure, and organization of ideas when writing in pairs. This may be attributed to the inherent characteristics of CW as it provides space for making suggestions, asking questions, and pooling ideas concerning the jointly written text. The findings, however, also imply that on average, students' performance concerning these measures has deteriorated on IW, particularly in the case of YS. Moreover, the findings reveal a difference between CAF and holistic measures. Although CAF measures indicate some degree of improvement in terms of complexity and accuracy on IW, the holistic measures show that students' performance on IW significantly dropped when compared with CW. This is all the more surprising given that students not only received the teacher's feedback but were already familiar with the task and working with one another. It also needs to be mentioned that the differences between holistic and CAF measures seem to support concerns raised by previous research that CAF measures may fail to grasp the improvement of learners' compositions which enhance with repetition (Hidalgo & Lázaro-Ibarrola, 2020). In accordance with Hidalgo and Lázaro-Ibarrola (2020), measures combining quantitative and holistic measures "to obtain a more thorough analysis of students' productions" (p. 514) have proven to be useful to get a more accurate picture of students' improvement.

## 6. Conclusion

The current study explored CW among M-A pairs on a common classroom task to gain a deeper understanding of CW in the M-A secondary school context. By exploring the difference between students' performance during CW and subsequent individual writing (IW) performance. On the whole, the findings point to limitations of M-A CW and the question arises as to whether the younger, particularly lower-level learners are likely to benefit from CW with an older partner. What is more, contrary to the claims of M-A supporters, the OS may not benefit either. In a similar vein, teachers should not assume that the OS students will as a matter of fact help their younger partners with complex aspects of writing including language use, structure, and organization of ideas. Therefore, if not organized well, this may hinder the younger and weaker students' learning and the expectations placed on OS may be too high to be met. This study has several limitations. Although more participants participated in the study, the complete set of data is available only from 28 students due to illness, attrition and an unexpected school event. Provided that the students in the current study were not

required to engage with the teachers' feedback, future studies can explore how and to what extent students' engagement with teachers' explicit feedback, text modeling, or reformulations provided to the text may affect CAF and holistic measures. Given that the CW process in this study was not at all structured and left alone to students how they approach the task, future studies may implement a clearly structured pair writing (see for example Yarrow & Topping, 2001) in which either member of the pair takes on the role of a helper (tutor) and a writer (tutee) which are identified at each step and which may include pairs generating ideas, producing a draft, reading their texts to each other, editing them, producing the best copy and finally evaluating. Given that this study focused solely on the exploration of students' pieces of writing, future research could explore a correlation between certain features of M-A peer interaction such as negotiation of meaning, LREs, patterns of interaction established by pairs on one hand, and the quality of students' writing on the other. For example, studies could highlight the relationship between the quality of writing and actual learning opportunities created during CW as measured by LREs. Finally, future studies need to give attention to learners' perspectives to understand their perceptions, goals, and feelings about their CW with a different age partner.

**References**

Adams, R., & Oliver, R. (2019). *Teaching through peer interaction*. New York: Routledge.

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, / other media interactions* (pp. 243-266). Ann Arbor: University of Michigan Press.

Adas, R. (2016). Focus on form: A critical review. *Language Teaching Research*, *20*(3), 405-428. http://doi.org/10.1177/1362168816628627

Alegr á de la Colina, A., & Garc á Mayo, M. P. (2007). Attention to form across collaborative tasks by low-proficiency learners in an EFL setting. In M. P. Garc á Mayo (Ed.), *Investigating tasks in formal language learning* (pp. 91-116). London: Multilingual Matters.

Azkarai, A., & Kopinska, M. (2020). Young EFL learners and collaborative writing: A study on patterns of interaction, engagement in LREs, and task motivation. *System, 94(*1). http://doi.org/10.1016/j.system.2020.102338.

Calzada, A., & Garc á Mayo, M. P. (2020a). L2 grammar learning through a collaborative writing task. In W. Suzuki & N. Storch (Eds.), *Languaging in language learning / teaching (*pp. 20-39). Amsterdam: John Benjamins.

Calzada, A., & García Mayo, M. P. (2021). Child learners' reflections about EFL grammar in a collaborative writing task: When form is not at odds with communication, *Language Awareness*, *30*(1), 1-16. http://doi.org/10.1080/09658416.2020.1751178

Calzada, A., & García Mayo, M. del P. (2020b). Child EFL learners' attitudes towards a collaborative writing task. *Language Teaching for Young Learners*, *2*(1), 52-72. http://doi.org/10.1075/ltyl.19008.cal

Chen, W. (2020). Disagreement in peer interaction: Its effect on learner task performance, *System*, *88*, 1-11. http://doi.org/10.1016/j.system.2019.102179

Coyle, Y., & Roca de Larios, J. (2014). Exploring the Role Played by Error Correction / Models on Children's Reported Noticing / Output Production in a L2 Writing Task. *Studies in Second Language Acquisition*, *36*(3), 451-485. http://doi.org/10.1017/S0272263113000612

Donato, R. (1994). Collective scaffolding in second language learning. In J. P. Lantolf, & G. Appel (Eds.), *Vygostkian approaches to second language research* (pp. 33-56). New Jersey: Ablex.

Duran, D., & Monereo, C. (2005). Styles and sequences of cooperative interaction in fixed and reciprocal peer tutoring. *Learning and Instruction*, *15*(3), 179-199. https://doi.org/10.1016/j.learninstruc.2005.04.002

Fortune, A. (2005). Learners' use of metalanguage in collaborative form-focused L2 output tasks. *Language Awareness, 14* (1), 21-38. http://doi.org/10.1080/09658410508668818.

Frank, S. S. (2014). Unterrichten in altersdurchmischten Klassen mit Young World 1-4. *Baar: Klett und Balmer*.

García Mayo, M. P. (2017). *Learning Foreign Languages in Primary School: Research Insights*. Bristol: Multilingual Matters.

García Mayo, M. P., & Agirre, A. I. (2016). Task repetition / its impact on EFL children's negotiation of meaning strategies and pair dynamics: an exploratory study. *The Language Learning Journal*, *44*(4), 451-466. http://doi.org/10.1080/09571736.2016.1185799

García Mayo, M. P., & Lázaro Ibarrola, L. (2015). Do children negotiate for meaning in task-based interaction? Evidence from CLIL / EFL settings. *System*: *An International Journal of Educational Technology / Applied Linguistics*, *54*(1), 40-54. http://doi.org/10.1016/j.system.2014.12.001

Gutiérrez, R., & Slavin, R.E. (1992). Achievement effects of the nongraded elementary school: A best evidence synthesis. *Review of Educational Research*, *62*(4), 333-376.

Heinzmann, S., Ries, S., & Wicki, W. (2015). Expertise Altersdurchmischter Fremdsprachenunterricht im Fach Englisch. *Zenodo*. http://doi. org/10.5281/zenodo, 2632679

Hernández, F., Larios, J., & Coyle, Y. (2017). Reformulation as a Problem-Solving Space for Young EFL Writers: A Longitudinal Study of Language Learning Strategies. In M. García Mayo (Ed.), *Learning Foreign Languages in Primary School: Research Insights* (pp. 193-222). Bristol, Blue Ridge Summit: Multilingual Matters. http://doi.org/10.21832/9781783098118-012.

Hidalgo, M. Ángeles, & Lázaro-Ibarrola, A. (2020). Task repetition / collaborative writing by EFL children: Beyond CAF measures. *Studies in Second Language Learning / Teaching*, *10*(3), 501-522. http://doi.org/10.14746/ssllt.2020.10.3.5

Hoffman, J. (2002). Flexible grouping strategies in the multiage classroom. *Theory Into Practice*, *41*(1), 47-52. http://doi.org/10.1207/s15430421tip4101_8.

Huff, C., & Raggl, A. (2015). Social orders and interactions among children in age mixed classes in primary schools - new perspectives from a synthesis of ethnographic data. *Ethnography and Education*, *10*(2), 230-241. http://doi.org/ 10.1080/17457823.2015.1017514

Katz, L. G., Evangelou, D., & Hartman, J. (1990). *The case for mixed-age grouping in early education*. Washington, DC: National Association for the Education of Young Children.

Kos, T. (2019a). Peer assistance among mixed-age pairs in mixed-age EFL secondary school classrooms in Germany. *European Journal of Applied Linguistics*, *7*(1), 61-112. https://doi.org/10.1515/eujal-2017-0013

Kos, T. (2019b). Patterns of Interaction: Analysis of Mixed-Age Peer Interactions in Secondary School Classrooms In Germany. *Journal of Language Teaching and Learning*, *9*(1), 1-29. https://www.jltl.com.tr/index.php/jltl/article/view/72

Kowal, M., & Swain, M. (1994). Using collaborative language production tasks to promote learners' language awareness. *Language Awareness*, *3*(2*)*, 73-93. http://doi.org/10.1080/09658416.1994.9959845

Leeser, M. (2004). Learner proficiency and focus on form during collaborative dialogue. *Language Teaching Research*, *8*(1), 55-81. http://doi.org/10.1191/1362168804lr134oa

Lindström, E., & Lindahl, E. (2011). The Effect of Mixed-Age Classes in Sweden. *Scandinavian Journal of Educational Research*, *55*(2), 121-144.

Little, A. W. (2001). Multigrade teaching: Towards an international research and policy agenda. *International Journal of Educational Development*, *21*, 481-497.

Lourenço, O. (2012). Piaget and Vygotsky: Many resemblances, and a crucial difference, *New Ideas in Psychology*, *30*(3), 281-295. http://doi.org/10.1016/j.newideapsych.2011.12.006.

McDonough, K., De Vleeschauwer, J., & Crawford, W. (2018). Comparing the quality of collaborative writing, collaborative prewriting, and individual texts in a Thai EFL context. *System*, *74*, 109-120. https://doi.org/10.1016/j.system.2018.02.010

Pape, M. (2016). *Didaktisches Handeln in jahrgangsheterogenen Grundschulklassen: eine qualitative Studie zur Inneren Differenzierung und zur Anleitung des Lernens* [Didactic action in heterogeneous primary school classes. A qualitative study of inner differentiation and guidance in learning]. Julius Klinkhardt.

Qin, J. (2008). The Effect of Processing Instruction and Dictogloss Tasks on Acquisition of the English Passive Voice. *Language Teaching Research*, *12*(1), 61-82. http://doi.org/10.1177/1362168807084494

Quail, A., & Smyth, E. (2014). Multigrade teaching and age composition of the class: The influence on academic and social outcomes among students. *Teaching and Teacher Education*, *43*, 80-90. https://doi.org/10.1016/j.tate.2014.06.004

Révész, A. (2011), Task Complexity, Focus on L2 Constructions, and Individual Differences: A Classroom-Based Study. *The Modern Language Journal*, *95*(1), 162-181. http://doi.org/10.1111/j.1540-4781.2011.01241.x

Robinson, D. R., Schofield, J. W., & Steers-Wentzell1, K. L. (2005). Peer and cross-age tutoring in math: Outcomes and their design implications. *Educational Psychology Review*, *17*, 327-362. https://doi.org/10.1007/s10648-005-8137-2

Ronksley-Pavia, M., Barton, G. M., & Pendergast, D. (2019). Multiage education: An exploration of advantages and disadvantages through a systematic review of the literature. *Australian Journal of Teacher Education (Online)*, *44*(5), 24-41. https://search.informit.org/doi/10.3316

Samuda V., & Bygate M. (2008). *Tasks in Second Language Learning.* New York: Palgrave Macmillan.

Saqlain, N. (2015). A comprehensive look at multi-age education. *Journal of Educational and Social Research*. http://doi.org/10.5901/jesr.2015.v5n2p285

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.

Smit, R., & Engeli, E. (2015). An empirical model of mixed-age teaching. *International journal of educational research*, *74*, 136-145. https://doi.org/10.1016/j.ijer.2015.05.004

Storch, N. (2005). Collaborative writing: Product, process, and students' reflections. *Journal of second language writing*, *14*(3), 153-173. https://doi.org/10.1016/j.jslw.2005.05.002

Storch, N. (2008). Metatalk in a pair work activity: Level of engagement and implications for language development. *Language Awareness*, *17*, 95-114. http://doi.org/10.1080/09658410802146644

Storch, N. (2011). Collaborative Writing in L2 Contexts: Processes, Outcomes, / 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. Manchón, & P. Matsuda (Eds.), *Handbook of Second / Foreign Language Writing* (pp. 387-406). Boston, MA/Berlin: De Gruyter Mouton.

Storch, N. (2021). Collaborative writing: Promoting languaging among language learners. In María del Pilar García Mayo (Eds.), *Working Collaboratively in Second/Foreign Language Learning* (pp. 13-34). Berlin, Boston: De Gruyter Mouton. http://doi.org/10.1515/9781501511318-002

Storch, N., & Wigglesworth, G. (2007). Writing tasks: The effects of collaboration. *Investigating tasks in formal language learning*, 157-177. https://doi.org/10.1177/0265532209104

Swain, M., & Lapkin, S. (2001). Focus on form through collaborative dialogue: Exploring task effects. In M. Bygate, P. Skehan, & M. Swain (Eds.), *Researching pedagogic tasks: Second language learning, teaching and testing* (pp. 99-118). Harlow: Pearson Education.

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

Tellier, A., & Roehr-Brackin, K. (2017). Raising children's metalinguistic awareness to enhance classroom second language learning. In M. P. Garcia Mayo (Ed.), *Learning foreign languages in primary school: Research insights* (pp. 22-48). Bristol: Multilingual Matters.

Thoren, K., & Brunner, M. (2019). State-wide implementation of mixed-age learning: which types can be identified, and do they differ in their school and teaching quality? *Zeitschrift für Erziehungswissenschaft*, *22*, 279-300. https://doi.org/10.1007/s11618-018-0841-z

Thurn, S. (2011). *Individualisierung ernst genommen. Englisch lernen in jahrgangsübergreifenden Gruppen (3/4/5)*. Bad Heilbrunn: Klinkhardt.

Topping, K. J., & Bryce, A. (2004). Cross-age peer tutoring of reading and thinking: Influence on thinking skills. *Educational Psychology*, *24*, 595-621. https://doi.org/10.1080/0144341042000262935

Villarreal, I., & Gil-Sarratea, N. (2020). The effect of collaborative writing in an EFL secondary setting. *Language Teaching Research*, *24*(6), 874-897. http://doi.org/10.1177/1362168819829017

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

Vygotsky, L. S. (1986). *Thought and language*. Cambridge, MA: MIT Press.

Wagener, M. (2014). *Gegenseitiges Helfen. Soziales Lernen im jahrgangsgemischten Unterricht* [Helping Each Other. Social Learning in Age-mixed Lessons]. Wiesbaden: Springer VS.

Wajnryb, R. (1990). *Grammar dictation*. Oxford: Oxford University Press.

Williams, J. N. (2012). Working memory and SLA. In S. Gass, & A. Mackey (Eds.), *The Routledge h/book of second language acquisition*. London: Routledge.

Yarrow, F., & Topping, K. J. (2001). Collaborative writing: The effects of metacognitive prompting and structured peer interaction. *British journal of educational psychology*, *71*(2), 261-282.

Young, A., & Tedick, D. J. (2016). Collaborative dialogue in a two-way Spanish/English immersion classroom. *Peer interaction and second language learning: Pedagogical potential and research agenda*, 135-160. https://doi.org/10.1093/applin/amu066