Dictogloss in the Primary School EFL classroom: Investigating the process, product and perceptions of collaborative writing

Author:

Asier CALZADA LIZARRAGA

Supervisor:

Prof. María del Pilar GARCÍA MAYO

2021

Acknowledgements

Carrying out a PhD thesis is usually depicted as a lonely process. There is undoubtedly a part of truth in that description, especially now that the whole research center building has become emptier as a consequence of the pandemic. Every morning that I step into the lab I miss meeting many of my old friends, who have either taken a different path in life or made the decision to go back to their hometowns. However, I wouldn't be writing a dissertation on the benefits of collaboration if I wasn't certain about the fact that the result of these long four years is not only the product of my own work but also of those who have accompanied me through this journey.

I would first like to especially thank my dissertation supervisor and mentor, Professor María del Pilar García Mayo. I can't imagine a more hard-working and proactive supervisor than Pilar. She has always been there to discuss my doubts, read my work, provide me with feedback in record time despite her many commitments and encourage me to publish ("never leave anything in the drawer!"). Pilar is also the one that sparked my interest in Second Language Acquisition while I was doing my master's, back in 2015, when I first learnt about the interactionist paradigm, and she made it sound fascinating. The story of how a graduate in Translation ended up researching tasks and young EFL children is in a great part due to her. I also really appreciated that Pilar got me involved in the MINECO project from the first minute, which has allowed me to develop my research skills enormously for the past four years. *Moitas grazas, Pilar!!!*

Being part of the *Language and Speech* (LASLAB) group at the University of the Basque Country (UPV/EHU) has also enabled me to grow professionally in a stimulating environment. Having so many experienced researchers as colleagues and working in the same field has clearly been a great bonus. Thanks also go to Professor Vicente Núñez Antón for the initial statistical advice and offering me timely guidance on the data analysis. LASLAB and the MINECO project (FFI2016-74950-P, 2016-2020) have offered me the means to attend countless conferences and training courses all over the world, where I've had the pleasure to meet other researchers and inspiring academics.

Besides, I'm thankful for the support from the Spanish Ministry of Universities, which granted me the FPU predoctoral contract (FPU16/01331), and thanks to which I have been able to work full time on my dissertation and gain some teaching experience at the UPV/EHU. I am equally grateful for having received (after a tediously long process) one of the short-time visiting scholarships to carry out a three-month research stay at the University of Leeds (United Kingdom). In one of the most uncertain times in recent history, Dr. Mohammad Ahmadian was tremendously kind to host me and offer his time and effort to discuss my work. He also introduced me to Centre for Language Education Research (CLER) and their IP Professor Maggie Kubanyiova, who invited me to share my research with all of them. Also, on the social side, this stay would have been much worse without the company of Marina, Carla, Alicia, Stephanie, Katie and George. Despite all the depressing covid-related news, you always managed to put a smile on my face and made the whole Leeds experience much warmer.

Data collection is always one of the most arduous tasks in Applied Linguistics research. In fact, my investigation would have been impossible without the participation of the two state schools where I conducted my study: LHI Abendaño and LHI Mariturri. Thanks to the school boards, teachers and parents for giving me access to school and making the whole procedure run so smoothly. And obviously, a big thank you to all the children that participated in the tasks, tests and interviews and showed so much involvement from the first to the last day.

The ups and downs that are inherently related to a PhD have been much easier to face thanks to the invaluable support of my colleagues, friends and family. First, thanks to the *predoc* team, María Luquin Urtasun and Janire Guerrero Gómez, for our nice chats in the office and cheering up my Monday mornings. I would also like to thank Olympia Simantiraki. I can't imagine a better colleague to sit next to. Be ready for my next visit in Crete! And just across the same corridor in the Micaela research center, one day I was lucky enough to meet at the right time the former FPU and already doctor Sergio López Sancio. Sergio has been one of the greatest supports during these PhD years. I will always cherish our *kafetxos* (aka coffee breaks), pizza and wine nights and conversations about absolutely everything. We didn't fix academia, but we definitely made it much more fun while were working together. Also, thanks to Aránzazu López Fernández and Dr. Luis Pastor Santamaría, I hope that by this time I'm already a worthy rival at *Fallera*! Although

I didn’t acquire any of your amazing cooking skills, you made the PhD experience much sweeter!

Outside the Micaela bubble, I must thank Daria Grigorieva, Даша, who at the start of my PhD was my Russian teacher and by the end has become one of my closest friends. Ты сделала жизнь в «деревне» гораздо менее скучной! Спасибо огромное за поддержку! I also wish to thank my old VTV friends, Josu, Miren, Yaiza C., Yaiza P. and Andrea. It’s always great to catch up with you and see how our lives are going after so many years! Thanks too to the new Vitorian girls, Anastasia Lavrenteva and Maiara Marques, who I’ve had the pleasure to share meals and daytrips with. And thanks to Ainara Cruz: despite the distance, we’ve always made it possible to meet up, be it in Spain or in England.

I can’t finish these acknowledgements without mentioning my family. Eskerrak eman nahi dizkiet nire gurasoei, Arantza eta Césarri, beti niregan izan duten konfindantzagatik eta eskaini didaten laguntzagatik. Argi daukat irakaskuntzan daukadan interesa zuei zor dizuedala! Mila esker nire aiton-amonei ere bai, zuengandik ikasi dudanagatik. Nahiz eta azken urtean ez garen asko ikusi, dei bakoitzean nire lanari buruz galdetu didazuenean izugarri poztu naiz. Eredu bat izango zarete beti niretzat!! Y muchas gracias a Diego, por estar siempre ahí y ser un apoyo fundamental. Hemos comprobado que dos tesis pueden caber perfectamente bajo el mismo techo. No dejes de recordarme nunca lo que verdaderamente importa en la vida.

Table of Contents

List of Abbreviations

3S	Third person singular present tense marker *-s*
3SAcc	Third person singular *-s* accuracy rate
Acc	Accuracy
Adq	Adequacy
AQ	Pretask attitude questionnaire
Borrow100	Borrowings per 100 words
C/T-unit	Clauses per T-unit
CAF	Complexity, accuracy and fluency
CALF	Syntactic complexity, accuracy, lexical complexity and fluency
CEFR	Common European Framework of Reference
CF	Corrective feedback
CLT	Communicative Language Teaching
Coher	Coherence
Cohes	Cohesion
Collab	Collaborative dictogloss group
CollaborationAtt	Attitudes towards collaboration
Comp	Comparison group (individual dictogloss)
Coord/T-unit	Coordinated clauses per T-unit
CW	Collaborative writing
DG	Dictogloss
DG1	Dictogloss on 3S "Naughty Laura"
DG2	Dictogloss on POSS "A celebration"
Disc	Discourse
DQ	Dictogloss task perception questionnaire
EEL	Early English Language
EFL	English as a Foreign Language
ESL	English as a Second Language
FFI	Focus on form instruction
FFI+Collab	Pretask FFI + Collaborative dictogloss group
FonF	Focus on Form
F-others	LREs related to other grammar forms
F-targets	LREs related to the target grammatical forms (3S and POSS)
FullCRatio	Full-fledged clause ratio
GI	Guiraud's Index
GJT	Grammaticality Judgment Task
GramErr100	Grammar errors per 100 words
GramErrFreeC	Grammar error-free clause ratio

HH	High-high proficiency pairing
HL	High-low proficiency pairing
HP	High proficiency
ID	Individual differences
ISLA	Instructed Second Language Acquisition
IU	Idea units
L1	First language
L1W	L1 writing
L2	Second/Foreign language
L2W	L2 writing
Lex	Lexis
LL	Low-low proficiency pairing
Ln	Any language learnt after the L2
LP	Low proficiency
LRE	Language-related Episode
LTH	Linguistic Threshold Hypothesis
ME	Metalinguistic Explanation
Mech	Mechanics
MLC	Mean length of clause
NoM	Negotiation of Meaning
ParaCRatio	Paraclause ratio
PKL	Previously known languages
POSS	Third person singular possessive dterminers his/her
POSSAcc	Third person singular possessive determiner accuracy rate
PreCRatio	Preclause ratio
ProtoCRatio	Protoclause ratio
RQ	Research question
SCT	Sociocultural Theory
SEM	Structural equation model
SFL	Spanish as a Foreign Language
SLA	Second Language Acquisition
SpellErr100	Spelling errors per 100 words
TBLT	Task-based Language Teaching
TR	Task repetition
USA	United States of America
WCF	Written corrective feedback
WritingAtt	Attitudes towards writing
YLs	Young learners
ZPD	Zone of Proximal Development

List of Tables

List of Figures

List of Appendices

Abstract

Research in Second Language Acquisition (SLA) has demonstrated that peer interaction has a facilitative role in second language (L2) learning (Philp et al., 2014; Sato & Ballinger, 2016). From a cognitive perspective (Long, 1983, 1996), interaction provides learners with opportunities to negotiate for meaning, make input comprehensible and modify their own output (Swain, 2005). According to sociocultural theories (Lantolf, 2012; Vygotsky, 1978), when peers interact, they can provide assistance to each other (scaffolding) and engage in metalinguistic discussions (languaging, Swain, 2006), which lead to the co-construction of new L2 knowledge. Task-based Language Teaching (TBLT) has long embraced the benefits of peer interaction, and scholars in this field have examined how these are mediated by task design, task implementation, as well as individual factors. Although low-proficiency young learners (YLs) have been underrepresented in TBLT (Newton & Le Diem Bui, 2020), there is a growing number of studies focusing on this population (Azkarai & Oliver, 2018; García Mayo, 2018b), a trend aligned with the earlier introduction of the L2 in primary school (Enever, 2018). The present dissertation aims to contribute to this body of research by investigating the impact of a collaborative writing task, the dictogloss task (Wajnryb, 1990), on young English as a Foreign Language (EFL) learners' attention to form and L2 writing, as mediated by task repetition (TR), pretask focus on form instruction (FFI) and individual variables, namely, attitudes and writing skills in the first language (L1).

Ninety-three (n = 93) child EFL learners (ages 11-12, L1 Spanish) performed a dictogloss task twice (Day 1 and Day 2). Learners were presented with two different dictogloss texts on each of the days, which targeted two problematic linguistic features: the third person singular *-s* (3S) and the possessive determiners *his/her* (POSS). Participants were assigned to three different task conditions: collaborative dictogloss (Collab, n = 28), pretask focus on form instruction on the target forms and collaborative dictogloss (FFI+Collab, n = 31) and individual dictogloss (Comp, n = 34). The oral interaction from the dyads working in collaboration was recorded and analyzed in terms of the quality and quantity of Language Related Episodes (LREs) they generated. In addition, all participants' reconstructed texts were assessed using text-based and rubric

measures. Moreover, in order to capture any changes in their individual L2 writing as a result of the task, all learners completed an individual written narration before and after the dictogloss intervention (using a pretest, posttest and delayed posttest). Besides, by means of questionnaires and focus group interviews, we tapped into these YLs' attitudes towards writing, collaboration and the task itself. Finally, we obtained an indication of their L1 writing skills by administering a written narration test in Spanish prior to the dictogloss intervention.

The findings show that the participants who performed the dictogloss task collaboratively mainly discussed mechanics (spelling and punctuation) and non-target grammar forms. Yet, the LREs on lexis were the longest and also the ones displaying more elaborate engagement from the dyads. The children made an extensive use of their previously known languages (PKL) for task resolution, and were able to correctly resolve most of their discussions. TR did not affect any of aspect of learners' attention to form, but FFI+Collab spent significantly less time on the task on the second day. Regarding attention to the target forms, in general the number of LREs was low, but receiving a pretask FFI helped children to produce significantly more episodes only in the case of 3S.

The patterns displayed by the dyads during the dictogloss task were mostly collaborative on both days, and the impact of TR on their behaviors was low. Contrary to our expectations, we failed to find any evidence to support the idea that collaborative work leads to better L2 writing performance than individual work. Nonetheless, we could observe a trend indicating that learners in the collaborative condition improved their grammar accuracy as a result of TR. Furthermore, coordination, which children used extensively in the first dictogloss enactment, was less present in their writing on Day 2. The impact of pretask FFI was reflected on a higher target form accuracy by learners in this group as compared to Collab and Comp. Finally, although the impact of the dictogloss task on learners' individual writing was low, the analysis of the type of clauses revealed that learners produced significantly fewer inaccurate and uncontextualized clauses after the pedagogical intervention, regardless of the condition in which they experienced it.

As to learners' individual variables, the YLs had a positive predisposition towards collaboration and writing, and the posttask questionnaire and interviews confirm their favorable opinion about the dictogloss task. Although learners do perceive the benefits of individual work, they are more inclined to pair work, and they also hold somewhat negative opinions about L2 writing as compared to L1 writing. What is more, we also found that attitudes played a more important role in explaining L2 writing achievement as compared to children's L1 writing expertise.

Altogether, these findings are encouraging for the use of the dictogloss task with YLs, as it has been proved to encourage attention to form and suit their affective needs. These results will be discussed in terms of their pedagogical implications and further lines of research will be suggested.

INTRODUCTION

For the last three decades, Task-based Language Teaching (TBLT) and Instructed Second Language Acquisition (ISLA) have sought ways of furthering learners' second language (L2) knowledge (R. Ellis et al., 2020). Aligned with the principles of Communicative Language Teaching (CLT) (Breen & Candlin, 1980), one of the practices that has been promoted by both ISLA and TBLT involves activities where learners work in collaboration. In fact, peer interaction is now widely considered to trigger behaviors associated with L2 acquisition (Sato & Ballinger, 2016). Research on task-based peer interaction has been primarily informed by two paradigms: the cognitive-interactionist and the sociocultural theories (R. Ellis, 2018). The former makes an emphasis on the cognitive mechanisms involved in interaction that are conducive to language learning, among others, the role of implicit and explicit learning (Long, 2015), determined in turn by the degree of attention and noticing (Schmidt, 1990, 2010). The latter, instead, drawing on Vygotskyan accounts of the human mind and learning (Vygotsky, 1978), argues that learning is a process continuously mediated by speech, and in L2 research, this verbalization of knowledge has been popularly known as *languaging* (W. Suzuki & Storch, 2020; Swain, 2006, 2010).

Within task-based peer interaction studies, a burgeoning field of research is the use of collaborative writing (CW) tasks (Storch, 2019a). CW consists in co-authoring a single written text, where all learners feel responsible for all stages of the writing task. This "shared ownership of the text produced" (Storch, 2016, p. 387) distinguishes CW from other modes of writing, such as peer review activities or group projects, where the work (searching for information, planning, writing, and reviewing) is usually divided among the learners, and hence, these activities are closer to cooperative writing. CW can still be considered a newcomer within TBLT, as researchers in this field have usually privileged oral tasks over the written modality (Byrnes & Manchón, 2014). Yet, there is now increasing evidence that writing can constitute an ideal site for L2 learning (Manchón & Vasylets, 2019). In fact, from a psycholinguistic perspective, while oral language is fleeting and is characterized by an online pressure component, writing, thanks to the permanent visual record and the availability of time, allows for a more self-regulated L2 output production, which in turn facilitates retrieval from long-term memory, as well as the detection of problems in the output (Gilabert et al., 2016; Williams, 2012; Zalbidea, 2021). In this regard, CW amalgamates elements that are fundamental for L2 learning and writing development, such as peer feedback, negotiations of form and meaning and L2 written output. In fact, the oral discussions that learners have during CW and their collaboratively written texts are an excellent tool to examine both noticing behaviors and L2 outcomes.

In a recent systematic review on face-to-face L2 CW research, Zhang and Plonsky (2020) identified several niches, including (i) the underrepresentation of young learners (YLs), (ii) the low number of learner-internal and learner-external moderating variables investigated, usually circumscribed to task mode and proficiency pairing, and (iii) the little attention paid to writing gains as a result of CW. We will now briefly discuss each of these areas in relation to the study presented in the current dissertation and justify why our investigation constitutes a novelty in the field.

Regarding YLs, the rise of early L2 learning in Europe and Asia over the last twenty years (Enever, 2018) contrasts with the scarcity of research conducted in this particular setting (García Mayo, 2017). In fact, far from being a problem exclusively related to CW research, several meta-analytic studies have found that YLs have been systematically neglected in studies on EFL writing (I. Lee, 2016), form-focused instruction (E. Y. Kang et al., 2019) or general ISLA research (Norris & Ortega, 2000). Furthermore, on some occasions the findings from very different contexts, such as immersion and early bilingual contexts, have been extrapolated and misinterpreted by policy-makers (Lyster, 2007). Nonetheless, early foreign language settings possess certain inherent characteristics (including a limited L2 exposure inside and outside school and the fact that all learners share the same L1) which make them very different from immersion settings. Therefore, specific research is required in this area to promote adequate policies and practices, and the investigation reported in this dissertation aims to shed more light on this topic by analyzing 11-12-year-old EFL children in the Basque Autonomous Community (BAC).

Secondly, regarding the type of moderating variables that impact CW performance, the current dissertation focuses on two underresearched learner-external variables, namely, task repetition (TR) and the provision of pretask focus-on-form instruction (FFI). As far as TR is concerned, it is expected that when learners perform a task for the first time, they will tend to focus on meaning, whereas the second (or subsequent) times they carry it out they will be more likely to focus on form (Long, 1996). Although most research on the impact of TR has used oral tasks (Ahmadian & Tavakoli, 2011; Bygate, 2009; García Mayo et al., 2018), some scholars have aimed to validate these findings in the written modality. For instance, in the case of individual writing, language performance has been reported to increase in terms of complexity, accuracy and fluency (CAF) (Amiryousefi, 2016) or at least in some of these dimensions (Khezrlou, 2020; Nitta & Baba, 2014). In the case of CW, the relatively few studies conducted to date report mixed results, both in terms of attention to form (Hidalgo & García Mayo, 2019) and written performance (Kim, Choi, et al., 2020; Kim, Kang, et al., 2020).

With regards FFI, there is now support for the idea that combining engaging tasks with timely focus on form (Long, 1991) can create a favorable context for L2 development (R. Ellis et al., 2020). Hence, one of the current debates revolves around when that FFI should be provided (i.e. the timing) (Khezrlou, 2019) and how explicit or implicit this instruction should be (Ghorbani, 2018). In the case of CW, the findings have been again varied, as some scholars report that FFI leads to more metalinguistic discussions (Kim & McDonough, 2011), whereas others do not report a significant influence in learners' attention to form (Leeser, 2004; Swain & Lapkin, 2001).

Finally, the systematic review by Zhang and Plonsky (2020) also indicated a shortage of studies investigating learner-internal variables (such as L2 motivation or their task orientation) and individual writing development as a result of CW. On the one hand, affective factors are believed to influence learners' interactional behaviors (Lambert, 2017; Sato, 2020b), and therefore, it is crucial to understand whether a certain task meets their affective needs. On the other hand, experimental designs involving individual pretests and posttests indicate a positive effect of CW in terms of rubric scores (Elabdali, 2021).

Therefore, the present dissertation seeks to contribute to the field of CW by addressing the aforementioned research gaps. Ninety-three (n = 93) L1 Spanish low-proficiency EFL children aged 11-12 were engaged in two dictogloss (DG) tasks (Wajnryb, 1990), in two consecutive weeks. Learners in the Spanish primary EFL context do not usually get much practice in L2 writing, but they are already experienced writers in their L1 by the end of elementary school (Mateo Cutillas, 2016; Núñez Cortés & Barrado Mendo, 2020). The choice of the DG task was motivated by the fact that it has been the most widely used task in CW research (Storch, 2019a; Zhang & Plonsky, 2020), and it has been proved effective for focusing adult (Alegría de la Colina & García Mayo, 2007; Storch, 2001), adolescent (Basterrechea & García Mayo, 2013; Kowal & Swain, 1994) and, more recently, child (Calzada & García Mayo, 2020a, 2021a) learners' attention to form. Moreover, it has also been considered an enjoyable task by child EFL (Calzada & García Mayo, 2020b; Kopinska & Azkarai, 2020) and ESL learners (Shak, 2006).

DG has been usually described as a focused task (Storch, 2016), whereby the researcher or the teacher targets a specific form that is a problematic linguistic feature. In the present study, we chose two forms that differed in their degree of salience and communicative value (Sato & Loewen, 2018): the third person present singular verb agreement morpheme *-s* (3S) (i.e. lower salience and communicative value) and the third person singular possessive determiners *his/her* (POSS) (i.e. higher salience and communicative value).

In order to investigate the impact of collaboration and FFI, these child learners were then divided into three task conditions: individual DG (Comp, n = 34), collaborative DG (Collab, n = 28) and pretask FFI + collaborative DG (FFI+Collab, n = 31). Participants from all conditions listened to a short passage twice, and during the second listening, took down notes of the key ideas individually. Then, they were required to reconstruct the text accurately keeping the content as similar as possible. While those in Comp reconstructed the text individually and then reflected in writing what difficulties they had encountered, those in Collab and FFI+Collab rewrote the text in dyads. Additionally, children in FFI+Collab carried out some form-focused activities on the target forms prior to the reconstruction. These data were then analyzed in terms of metalinguistic reflections. Besides quantifying the instances of languaging as Language Related

Episodes (LREs) (Swain & Lapkin, 1998), we were equally interested in determining how children resolved these discussions, how deeply engaged they were (McDonough & Hernández González, 2019) and what language they preferred for collaborating (Martínez Adrián & Arratibel Irazusta, 2020).

Another aim of the present study involved assessing the quality of collaborative work during DG. In fact, dyadic patterns have attracted a great deal of attention from scholars investigating collaborative tasks with adults (Kim & McDonough, 2008; Storch & Aldosari, 2013; Watanabe & Swain, 2007) and YLs (Azkarai et al., 2019; García Mayo & Imaz Agirre, 2019; Oliver, 2002), and these behaviors have been suggested to influence learners' attention to form (Chen, 2018; but see Azkarai & Kopinska, 2020). Besides, in this dissertation we looked into the quality of the writing produced as a result of the DG task, and more specifically, we examined whether there were any differences between the individual and the two collaborative conditions (Collab and FFI+Collab). In this sense, previous research has generally found an advantage of the collaborative mode over the individual in terms of accuracy (Elabdali, 2021). Moreover, both the dyadic patterns and the reconstructed texts were analyzed in relationship to the variable of TR.

Whether experiencing the DG task in any of its conditions (Comp, Collab or FFI+Collab) generated changes in these learners' individual writing was another research gap we intended to fill. In order to do so, the participants carried out an L2 writing pretest, posttest and delayed posttest (one week before, one week after and two weeks after the DG intervention, respectively). Both in the case of the DG reconstructions and the L2 writing tests, the complexity and accuracy of child learners' production was assessed using a range of traditional text-based measures (e.g clauses/t-units or number of errors per 100 words) (Polio & Friedman, 2016), which are well-founded within the CAF framework (Housen et al., 2012). Furthermore, taking the characteristics of young low proficiency learners in consideration, we also employed a range of specific measures, such a clause classification combining impressionistic and text-based criteria (Torras, 2005), the amount of coordination (Bardovi-Harlig, 1992; Mylläri, 2020), as well as an analytic rubric for evaluating EFL narrative skills (Government of Navarre, 2017).

Last but not least, the present dissertation also seeks to contribute to the study of affective dimensions in relation to CW in the early EFL classroom. More specifically, we investigated three dimensions related to the CW learning experience (Dörnyei, 2019): (i) child learners' attitudes towards writing and collaboration, (ii) their DG perceptions and (iii) the contribution of children's attitudes towards writing and their L1 writing expertise to their L2 writing achievement. These affective variables were tapped using two context-specific questionnaires: the Attitude Questionnaire (AQ) and the Dictogloss Questionnaire (DQ). In addition, the learners' responses to these questionnaires were supplemented with focus group interviews. The inclusion of the predictor variable of L1 writing responds in turn to the need to shed light on a question which has received so far little attention from L2 writing studies, namely, the extent to which learners' L1 writing skills are transferable to the L2 writing process (Manchón & Williams, 2016; Schoonen et al., 2011). In order to examine this potential correlation between the written product in Spanish and English, child learners were required to complete a story continuation task in their L1 prior to the DG procedure, which was then assessed using an analytic rubric (Fernández et al., 2019).

The main findings of the present study reveal that YLs who performed the DG collaboratively attended more to mechanics (spelling and punctuation) and non-target grammatical features, and to a lesser extent to lexis, target forms (3S and POSS) and discourse. Moreover, they resolved the vast majority of their metalinguistic discussions correctly relying only on their own resources. On the other hand, participants resorted extensively to Previously Known Languages (PKL) during their collaborative dialogue, especially when discussing lexis, the category in which children showed more elaborate engagement. As far as learners in the individual condition are concerned, they wrote few metalinguistic reflections after reconstructing the text, and these were mainly devoted to spelling and lexis. Regarding the impact of TR, this variable did not influence learners' attention to form, but it did lead to a significant decrease in task time in one of the conditions (FFI+Collab). Moreover, adding a pretask FFI to the DG task appeared to significantly influence the amount of LREs on 3S only.

Regarding dyadic patterns, the collaborative behavior predominated on both task days and the influence of TR was again low. However, pairs in FFI+Collab also showed some indication of more uncollaborative dynamics than their counterparts in Collab. Besides, the analysis of DG text reconstructions did not show an advantage of the collaborative mode over the individual one, but adding a second encounter allowed us to observe that learners in Collab and FFI+Collab increased their grammatical accuracy, while those in Comp kept their scores at the same level. However, irrespective of task condition, learners on Day 2 used significantly less coordination in their reconstructions. The pretask FFI also seemed to have positively influenced target form accuracy, as FFI+Collab scored the highest. Regarding the influence of the DG task on individual L2 writing development, we could not find enough evidence to support our initial claim, but there was a trend for all learners to produce significantly fewer inaccurate and uncontextualized clauses after the DG intervention.

As to learner-internal variables, there was a positive predisposition from children to write and collaborate, but they also had a more unfavorable view of EFL than L1 writing. Furthermore, our results showed that these YLs generally enjoyed the DG task, especially in the case of learners who performed it in collaboration without a pretask FFI stage. To conclude, concerning the relationship between L2 writing achievement and the selected predictor variables (writing attitudes and L1 writing expertise), our model indicated a heavier influence of children's affective factor as compared to their command of narration in Spanish.

The rest of this dissertation is organized in eight different chapters. Part I (**Literature Review**) comprises four chapters that provide the necessary background to the empirical study reported in Part II **(The Study)**, which also consists of four chapters. Chapter 1, **Child foreign language learning**, describes the main characteristics surrounding early foreign language learning, emphasizing three aspects that make this context different from naturalistic settings, namely, the age of onset, time of instruction and the learning mechanisms involved. Then, we summarize the main reasons that account for the popularity of early English learning programs across the world, and specifically address the context of primary EFL in Spain and the BAC. Chapter 2, **Task-based Language Teaching and Learning**, provides the rationale for implementing this innovative framework in the L2 classroom and reviews the main theoretical underpinnings from the three main strands that inform this area of research (the interactionist, the sociocultural and the psychological perspective). Chapter 3, **Task-based Language Teaching and Learning with children**, aims to bring together the two previous chapters by, first, providing justification as to why TBLT can be a suitable approach for YLs, and secondly, by summarizing the research on child task-based interaction in second language and foreign language settings, which has substantially increased over the last years.

In Chapter 4, **Collaborative writing tasks**, we concentrate on this specific task mode, which combines the oral and written modality and is the object of study of the present dissertation. We will begin by highlighting the learning affordances of L2 writing (from the so-called *writing-to-learn* perspective), and move on to argue, from a sociocultural paradigm, why adding collaboration to L2 writing tasks can further support L2 learning. After reviewing these theoretical aspects, Chapter 4 presents the most important empirical findings in CW research that compared individual and collaborative writing performance. The next sections in this chapter are devoted to the variables that impact CW performance, and they are grouped in two categories: task-related factors and individual variables. Among the former, we describe the influence of task type (including a thorough review of studies that have previously used collaborative DG), task repetition and pretask FFI. Regarding the latter, we present previous findings on the impact of L2 proficiency, affective factors and L1 writing skills. Finally, Chapter 4 concludes with a section addressing the assessment of L2 written production from a CAF perspective.

Part II **(The Study)** contains the empirical research. Chapter 5, **The study**, provides the motivation, the research questions and the hypotheses guiding our study. The ten research questions we entertained have been classified into three main sets, which are summarized in the following topics: (i) DG, focus on form and the impact of task-related variables, (ii) the impact of collaborative work, and (iii) individual differences. Besides, the information about the participants, the setting, the materials and the procedure followed is presented. Within materials, the choice of the DG target features (3S and POSS) is justified. Chapter 5 concludes with a detailed description of the codification and data analysis. Chapter 6, **Results**, presents the results obtained for each of the research question sets, and these are subsequently discussed in relation to the initial hypotheses in Chapter 7, **Discussion**. Finally, in Chapter 8, **Conclusions**, some overall conclusions are drawn, the implications for pedagogy are discussed, the limitations of the current study are acknowledged and further lines of research are suggested.

PART I: Literature review

CHAPTER 1 Child foreign language learning

1.1 Children in language acquisition studies

Children have been the object of study for language acquisition scholars since the inception of the field (Brown, 1973). In the case of first language (L1) acquisition studies, scientists have always wondered how the language faculty could develop so quickly in such precarious circumstances, among others, a still developing neurological apparatus and the enormous variability in the speech input that children rely on (Kuhl, 2010). This ease to acquire the mother tongue led to the theoretical assumption that all humans are endowed with an innate capacity for language, which enables us to set the right parameters through hypothesis testing during the first life stage (Chomsky, 1986). In other words, as VanPatten and Williams explain (2014), "children come to the task of language already knowing a great deal; they simply need the triggering data in the input for language acquisition to take place" (p. 24).

In the field of a second language acquisition (SLA), a great number of scholars have wondered if that disposition still holds true. Many of the studies from the 70s and 80s were devoted to analyzing how children acquired a second language (L2) while they completed the acquisition of an L1 or right after they had completed it, as in the case of children whose parents had migrated to a different country. For example, in the study by Johnson and Newport (1989), the authors examined the acquisition of English by a group of Chinese and Korean speakers (n = 46, ages 3-39, length of residence 3-26) who had emigrated to the United States of America. The authors concluded that it was the age of arrival that determined the participants' L2 morphosyntactic proficiency (measured by means of a grammaticality judgment task), since the findings revealed that target-like command started to decline when participants had arrived at the age of 7 and it was lost at around 15-17 years.

The underlying idea of studies like the one by Johnson and Newport was to prove a potential advantage of children over adults. Thanks to a greater knowledge of their L1 and a more developed analytic capacity, at first adults seemed to grasp the L2 more easily. In other words, the initial rate (Krashen et al., 1979) was higher for adults, but after a longer period of residence the children always overtook their older counterparts. That is, the level of ultimate attainment was always higher the younger they started to acquire the language.

These studies were also motivated by Lenneberg's (1967) Critical Period Hypothesis (CPH), according to which there is a certain period of time when children are biologically predisposed to have a full acquisition of language. That period, in fact, would go from the age of two to somewhere around puberty. Lenneberg's hypothesis have heavily influenced not only L1 acquisition abut also SLA studies (for a review, see DeKeyser & Larson-Hall, 2005). However, the study of other variables influencing L2 acquisition has downplayed the role of maturational constraints when it comes to certain linguistic aspects, such as ultimate attainment in morphosyntax. For instance, in a replication of Johnson and Newport's study carried out by DeKeyser (2000) with Hungarian-speaking learners of English in the USA (n = 57), those participants who were late arrivals and scored within the range of child arrivals (or came close to it) were considered high-aptitude participants. In other words, aptitude, understood by DeKeyser as the verbal ability in the L1, plays a determining role in adult L2 acquisition. These findings offered support for Bley-Vroman's (1989) Fundamental Difference Hypothesis, which posits that there is a qualitative difference between how children acquire their L1 and how adult learners acquire an L2. Following this hypothesis, children primarily rely on implicit learning mechanisms (i.e. picking up certain aspects of language from mere exposure to input) whilst adults learn explicitly (DeKeyser, 2003), that is, involving conscious awareness from their part.

Nonetheless, the views on the role of input in child L2 acquisition have been very much refined since then (DeKeyser, 2020). For example, the results of the study conducted by Granena (2014) in an immigrant context in Spain (n = 50, L1 Chinese, L2 Spanish, age of onset 3-6) suggest that individual differences (IDs) like aptitude do play a role in certain areas of the acquisition of morphosyntax (such as inflectional morphology) by early childhood learners. Although all the studies reviewed so far correspond to naturalistic settings, in early instructional contexts, where input is not plentiful or might not be of a of good quality, "implicit learning works less well and high aptitude for explicit learning becomes a necessary condition for reaching near-native competence" (DeKeyser, 2020, p. 80). In the remainder of this section, we will concentrate on the characteristics of early foreign language (FL) contexts, as this is the context in which the present dissertation was set.

1.2 Factors influencing early FL learning

There are clear differences between naturalistic and FL contexts which simply cannot be disregarded. Muñoz (2008) summarized the main characteristics of FL settings, which include the following: (i) learners have a limited exposure to the L2 and its main source is the teacher, (ii) the L2 is not the learners' language of communication, (iii) the teacher's quality of the L2 may be limited and non-target-like, and (iv) the L2 is not spoken outside school. However, more recent descriptions of early FL settings (Azkarai & Oliver, 2018) already acknowledge that there is now a greater access to the L2 outside school thanks to online resources, especially in the case of a global language like English, whose role in today's world in relation to early foreign language policies will be discussed below. Even so, more exposure to the language is not the same as having the possibility of interacting in the L2. We will now discuss some of the variables that have been claimed to influence the acquisition of the L2 in early FL contexts, namely, the age of onset, exposure time and learning mechanisms.

1.2.1 Age of onset

At the beginning of the 21st century, two main research projects were carried out in order to clarify the role of the starting age in an English as a foreign language (EFL) setting: on the one hand, the studies conducted by the Language and Speech research group in the Basque Autonomous Community (BAC) (García Mayo & García Lecumberri, 2003); and, on the other hand, the Barcelona Age Factor (BAF) project (1995-2002), led by Muñoz (2006) in Catalonia. What these authors concluded from their studies was that the advantage of early starters over late starters was not maintained. Older learners' higher learning rate (that is, better learning efficiency) translated into higher scores in the different tests (cloze test, dictation, grammar test, listening comprehension and written composition), which were administered at three different times. At most, in the less cognitively demanding tasks, differences between younger and older starters became non-significant or were reduced. However, in the more cognitively-demanding ones the older learners' superiority persisted. Among the reasons that explain this advantage over young starters, the researchers pointed to older learners' cognitive maturity, and suggested that other factors, such as quality and quantity of the input, played a more important role than the starting age.

In the specific case of writing, the linguistic competence which is most central for the current dissertation, two studies should be mentioned. First, Lasagabaster and Doiz (2003) examined the written production of three groups of EFL learners in the BAC, who differed in their L2 starting age (Group A: age 4/5, Group B: age 8/9, and Group C: AGE 11/12). At the time of the study, they were respectively 11-12, 15-16 and 17-18 years old. They all had to compose a letter addressed to an English host family, and this production was assessed in terms of a holistic scale and error frequency analysis. Their findings indicated maturational constraints in the domain of writing, as those learners who were older and had had more writing practice performed better than 11-12-year-old learners.

The second study was conducted by Torras et al. (2006) and it compared the written production of two Catalan EFL learner groups with different ages of onset during their childhood: group A (age of onset = 8) and group B (age of onset = 11) at three different testing times (T1, after 200 hours of instruction; T2, after 416 hours and T3, after 726 hours). The learners had to carry out an introductory writing about themselves, which was analyzed using different complexity, accuracy and fluency (CAF) measures (a framework that will be reviewed in more detail below). The analysis showed that there were no significant differences between the early and the late starters in ten out of fourteen measures at the three different testing times, and where that was the case, it was always in the form of an advantage of group B over group A. Therefore, they suggested that in the case of writing an early start does not seem to guarantee a better level of attainment. What is more, the authors explained that there is a turning point at around the age of 12, where grammatical complexity measures tend to increase. In fact, the early starters in this group received 210 hours less of instruction after that age than the late starters, which could explain part of the advantage of group B over group A. Thus, the authors wondered if the same advantage could be shown in the case that an early start in the L2 were accompanied with more instruction over a longer period of time. Last but not least, they also acknowledged the potential effect of teaching methods, as before the age of 12 children do not usually receive so much grammar or focus on form (FonF) instruction, and neither do they work as much on writing skills (oral communication prevails). Hence, different pedagogical practices could also impact the attainment in the foreign language writing competence to a greater extent than the age of onset.

1.2.2 Time of exposure or instruction hours

Regarding instruction hours, Muñoz (2008) draws a parallelism between studies looking at this variable in FL settings and those looking at length of exposure or length of residence in naturalistic settings. Although research from naturalistic settings informs that length of residence is no longer a predictor of L2 proficiency after the initial period, that is, after the early ages of a learner (Long, 2007), in instructional settings this lack of interrelation does not hold true. In fact, the general scarcity of input in FL settings (reduced to three to seven hours at most per week) makes the continuous exposure well beyond the so-called initial period incredibly valuable in the L2 acquisition process. As Muñoz (2008) explains, the amount of language exposure that a FL learner would need to equal a naturalistic learner would largely exceed a lifetime (more than 200 years), not to mention that a FL learner is deprived of a great of number of linguistic constructions and speech acts that are not present in the classroom.

Yet, empirical research on how beneficial increasing the number of L2 instruction hours is in an FL context has shown mixed findings. An example of a branch of studies that has aimed to examine the impact of teaching hours has been the Content and Language Integrated Learning (CLIL) studies, which compared this setting to mainstream EFL ones. In fact, CLIL has been one of the ways that many countries have sought to compensate for the L2 input shortage. This methodology involves using an L2 (foreign, regional or minority languages) as a medium of instruction, in addition to the mainstream L2 lessons. In CLIL, not only is exposure increased, but also learners are provided with meaningful input that, eventually, leads to the processing of meaning (Muñoz, 2007).

Most CLIL studies have taken place in secondary school settings (learners ranging from 12 to 18 years old), as this is the usual time when such programs are initiated (Martínez Adrián & Gutiérrez Mangado, 2015). However, studies comparing CLIL vs non-CLIL (or mainstream EFL) in this age range have failed to find significant differences in specific areas of language acquisition. For example, García Mayo and Villarreal Olaizola (2011), examined the omission rate of English suppletive and affixal morphology at two testing times in secondary school, when learners were 14-15 years (Time 1) and when they were 17-18 (n = 78 in total). The hours of exposure between both groups (CLIL and mainstream EFL) varied considerably: the former had 875-910 hours of exposure at Time 1, and 1443 hours at Time 2, whereas the latter had, respectively, 693 and 990 hours. Yet, the authors could not find any significant difference in omission percentages except for two linguistic features (*be* auxiliary and auxiliary and copula *be*) at Time 1. It is suggested that CLIL could have a "ceiling effect with the morphemes reported in this study" (p. 143), and that even an increase in exposure hours is not translated into a learning or uptake increase.

Besides the acquisition of morphosyntax, other areas of language, such as the writing competence, have been investigated with regards to hours of instruction provided by CLIL and non-CLIL programs. Lasagabaster (2008) compared the written production (a letter to a host family) of two CLIL groups and a non-CLIL one in Spain. The first two corresponded to two different years of secondary education in which the younger participants were 14-15 (n = 57) and the older 15-16 (n = 28). The non-CLIL group were the same age as the older CLIL ones (n = 113). Those who attended mainstream EFL received 3 hours per week, whereas the CLIL learners received 4 hours per week. In the same-age comparison, CLIL learners performed significantly better than non-CLIL ones in all writing domains (assessed through a holistic scale). What is more, younger CLIL learners performed significantly better than their older non-CLIL counterparts did. In this kind of studies, however, a question lies as to the extent to which the advantage of CLIL learners comes from a larger amount of teaching hours or, rather, the CLIL methodology per se (Martínez Adrián & Gutiérrez Mangado, 2015), which tends to focus on meaning and communication to a greater extent than mainstream EFL (Lasagabaster, 2008).

In order to overcome this limitation, Tragant et al. (2016), this time focusing on the child EFL population, compared CLIL vs mainstream EFL instruction by controlling for the variable of instruction time, as it involved a within-subjects design. In their small-scale study, the authors wanted to find evidence of any advantage of the CLIL methodology when it comes to L2 vocabulary acquisition. In the fall term, the twenty-two leaners (n= 22, ages 8-9, L1 Spanish and Catalan, L2 English) worked on six units from a mainstream EFL manual, and in the winter term, they went through five units of a Science module which was delivered in English. The respective vocabulary gains were tested before and after each of the terms. Results showed that students made significant vocabulary gains in both learning contexts, however the comparison of those gains resulted in significantly higher mean gains in the EFL than in the CLIL setting (with a large effect size d = 0.50). The authors suggest that this difference in favor of mainstream EFL could be due to two reasons: on the one hand, a more frequent use of the L1 in CLIL, and on the other, exposure to more relevant vocabulary in EFL for young learners' lives.

To summarize, the inconclusive results derived from the studies examining the impact of time of exposure in FL settings need to be interpreted in parallel with the influence of the type of instruction learners receive. In fact, mainstream EFL is characterized by a more "formal" approach to language teaching, as opposed to the communicative one embraced by the CLIL methodology. This is related to an ongoing debate in cognitively-oriented SLA concerning the learning mechanisms (implicit and explicit) of young L2 learners (Muñoz, 2008), as discussed below.

1.2.3 Learning mechanisms

As in the case of L1 acquisition and L2 acquisition in naturalistic settings, it has been suggested that cognitive factors limit, to a certain extent, the way children learn an L2 in an FL setting. According to the traditional perspective, based on the comparison of child L1 and adult L2 learning, children naturally rely on implicit mechanisms, whereas in adults these are reduced, and thus, they have to draw on explicit resources (Bley-Vroman, 1988). Consequently, just as children can perceive phonetic, lexical and morphosyntactic cues in the L1 input and build an internal linguistic system by deducing the rules subconsciously, SLA scholars originally contended that children acquire the L2 in the same way and dismissed the influence of formal teaching to some extent (Krashen, 1983; Newmark, 1966).

Nonetheless, more recently this view has been confronted with what actually occurs in FL settings. As explained above, child learners' exposure in instructional settings is very limited and, therefore, it is not likely that they can build a linguistic system based only on exposure. Furthermore, in the last stages of childhood (or early adolescence), that is, the period we are concentrating on in the present dissertation, learners have been shown to be more cognitively, linguistically and socially ready for explicit instruction and language analysis (Philp et al., 2008). This development is outlined in Piaget's stages of development (1926, 1955), according to which in Stage 4 (ages 11-12), in the so-called formal operational stage, children have a rational, systematic and abstract thinking, similar to a scientific one (Pinter, 2011). However, further criticisms of Piaget's theory and experiments argue that reaching such level does not necessarily mean that children will always act according to the rules of formal logic (Wood, 1998).

From an information processing perspective, children's attentional strategies also improve as they grow older, particularly after middle childhood (Strutt et al., 1975), as well as their processing capacity (Kail & Salthouse, 1994) and memory store, both showing a peak rise after the age of 11-12 (Brunswick et al., 1932). Considering all the described cognitive specificities, we could conclude that older children are equipped to perform more complex and challenging tasks while learning an L2. In other words, tasks which do not primarily rely on implicit learning, but which actually engage learners in explicit form-meaning mapping connections.

In this vein, recent research shows how children in FL instructional settings can benefit from a more explicit approach. Tellier and Roehr-Brackin (2017) aimed to show the influence of children's metalinguistic awareness on FL learning and, in particular, whether focus on form instruction (FFI) (Spada, 1997) is as beneficial for this population as it has been demonstrated for adults (Kang & Lee, 2019; Norris & Ortega, 2000). In their study, metalinguistic awareness is understood as the combination of grammatical sensitivity and inductive language learning aptitude (Skehan, 2002), which are part of a wider and multicomponential construct called language learning aptitude

(Abrahamsson & Hyltenstam, 2008). The study consisted of two parts. For the first one, a total of 178 L1 English learners, aged 8-9, were divided into different treatment groups: Esperanto plus dedicated focus on form activities (E+) (n = 53), only Esperanto, (E) (n = 45), German (G) (n = 39) and Italian (I) (n = 41). Children were tested on a metalinguistic awareness test, and the results showed that E+ performed significantly better in the post-test than G and I. In the second part of the study, Tellier and Roehr-Brackin wanted to compare the role of these “starter” languages in the learning of French, which was part of the children’s school curriculum. This way, they could see if having one language which was constructed and more transparent, such as Esperanto, could help learners establish crosslinguistic inferences. All learners attended several lessons of French with a dedicated FFI. The statistical analyses showed that the starter language did not play a role in the French test gains, but the children’s metalinguistic awareness was correlated with their French results. Therefore, these findings suggest that FFI can actually play a role in the early learning of a FL, where exposure is minimal, and also that children do not only rely on implicit mechanisms, but rather that they draw on their explicit linguistic ability.

In a subsequent study, Roehr-Brackin and Tellier (2019) investigated more in detail the role of language learning aptitude and language analytic awareness with the same sample of learners (n = 111). They employed the MLAT-E(UK) tests, which comprises four subtests assessing memory and language-analytic abilities. Firstly, they concluded that, in line with previous research, both language learning aptitude and language analytic ability fall under the same construct (i.e. showed the same variance of response, as indicated by the factor analysis). However, they also suggest that it is difficult to sustain that this aptitude is innate (Ranta, 2002), as their data demonstrated that it was dynamic. Yet, it still needs to be shown whether the dynamicity is due to general cognitive development or the type of instruction (learning experience) they receive. Secondly, the authors found a strong correlation between the children’s MLAT-E scores at the beginning of the experiment and the following L2 French improvement in grammar, reading and listening. Interestingly, however, writing scores did not correlate with overall aptitude, indicating the influence of other potential predictors on this domain (such as literacy skills in the L1). The overall variance explained in proficiency scores explained by MLAT-E was $r = .56$.

Finally, according to Roehr-Brackin and Tellier (2019), the fact that more language-analytic related subtests have a greater predicting value than memory related subtests in L2 gains makes young learners (YLs) more similar to adult L2 learning patterns than what is generally thought. Hence, young L2 learners would not be so reliant on memory and implicit mechanisms "as the type of instruction to which young learners are exposed may be just as important as chronological age in determining the use of primarily implicit versus primarily explicit learning mechanisms" (p. 1127).

To conclude, these recent studies suggest that there is room for FFI in early stages of L2 learning. Favoring explicit mechanisms which encourage language-analytic abilities among YLs can actually facilitate a greater L2 success or speed up their acquisition process in settings where L2 input is limited, such as FL instructional contexts. As we will explore in the next section, early EFL programs are increasingly popular, hence the need to test suitable teaching methodologies through specific research devoted to this population.

1.3 English as a global language: early EFL language policies

1.3.1 An international perspective

Although foreign languages have always been taught and learnt in early stages of life, it was not until the 1970s, and especially, until the turn of the century that their instruction began to take off across all social strata. In fact, what had previously been a luxury became available to middle and lower class by means of democratic L2 teaching programs established in the education curriculum of a high number of western countries (Enever, 2018). There are different factors which explain this popularization, among them (i) the global trade, (ii) the economic influence and hegemony of USA, and (iii) the political and war conflicts in different parts of the world which have led to migrations from poorer to richer countries, and consequently, to an enormous social diversity arisen in many societies, such as in Europe.

As Enever (2018) explains, in Europe and Asia the preferred foreign language since the beginning of the 21st century has been English. In fact, it is estimated that nowadays half a billion children over the world are learning English. In the case of Europe, the latest Eurydice report (European Commission/EACEA/Eurydice, 2017) shows that more learners in primary education are learning a FL than a decade ago, as for “2014, at EU level, 83.8% of all students attending primary education studied at least one foreign language” (p. 11), compared to 67.3% in 2005. Furthermore, in 2014, according to the same report, 79.4% of learners in primary education studied English, 18.7% more than in 2005. The lowering of the starting age clearly contributed to this rise.

There are different reasons that account for an increase in Early English Language (EEL) study programs. Together with the economic changes mentioned (the USA supremacy, a more a global trade as a result of the opening of the Asian region to the capitalist markets, the fall of the Soviet Union in the 90s and migration flows), we must note the popularity of digital technologies, which have made the world more interconnected, as well as the access to travel and the increased lifespan. Enever (2018) points out that most parents perceive English as necessary for their children to participate in contemporary society and to compete globally. It is no wonder that nowadays English is perceived as a basic skill along with literacy and numeracy.

National education authorities, aware of such needs, have been permeable to the recommendations of international institutions (World Trade Organization, Organization for Economic Co-operation and Development), and hence, they have taken measures to make English available from an early start. However, these international policies have sometimes been implemented with a “short-termist” approach (Enever, 2018, p. 24). Moreover, SLA research has sometimes been misinterpreted, overgeneralizing findings from immersion settings (e.g. Canadian French immersion programs, Swain & Lapkin, 1991) and early bilingual acquisition (e.g. Spanish and English bilinguals in the USA, Bialystok, 2001). As explained in the previous section, there are multiple reasons why FL settings differ from the aforementioned ones (the quality and quantity of the input or the function of the L2 outside school, to mention a few), and in fact, “the earlier the better” is not always sustained by empirical findings.

Another consequence of these global EEL measures has been an education standardization, the development of knowledge measurements, and a series of policy and pedagogical borrowings (Enever, 2018). A clear example of these processes is the PISA (Program for International Student Assessment) transnational test, or in the case of foreign languages, the European Survey on Language Competencies (Costa & Albergaria-Almeida, 2015). Nonetheless, data from these tests have sometimes been hastily analyzed, without taking into consideration other specificities of the countries, such as their Gross Domestic Product (GDP), population size or teaching traditions. In addition, the uneven growth of English learning resources has sometimes led to inequalities in the provision of a quality instruction in the public educational system, which in certain countries has been supplemented by attending extracurricular English lessons (being especially popular in countries like Greece and South Korea) to ensure a more successful academic performance. In other words, the goal of equity in the provision of English has not been achieved and has been shown to be dependent on the families' socioeconomic status.

As far as language competence testing is concerned, both public and private institutions have placed a greater emphasis on developing appropriate tools for assessing YLs. In a context where extracurricular English lessons have become the norm, these measurements also aim at satisfying the parents' need to know how their children are progressing in the L2 (this is what Enever circumscribes within the concept of "parentocracy" in language policy formation, p. 25). For instance, Cambridge Assessment English (n.d.-b) has developed a range of English tests (*Starters, Movers and Flyers*) adapted to children, which go from a pre-A1 to an A2 level, that is, an elementary proficiency according to the Common European Framework of Reference (CEFR) (Council of Europe, 2001). Regarding the efforts made by public institutions, we must refer to the update of the CEFR in the recently published Companion Volume (Council of Europe, 2018). Indeed, one of the criticisms raised against the original CEFR was the fact that it was based on adults' assessment assumptions of child learners (Butler, 2019). In fact, sometimes the ideas behind standardized testing may clash with children's inherent trait, as there are "substantial individual differences in cognitive, sociocognitive, and linguistic developments among children within the same age groups" (Butler, 2019, p. 481). Yet, the efforts to adapt the CEFR have been reduced to creating a pre-A1 level and further-dividing the A1 and A2 stages, where descriptors have been reworded for two distinct groups of learners, namely, those aged 7-10 and those aged 11-15. Some specific guidelines for teaching foreign languages, and namely English, to YLs have remained out of the Companion Volume, as well as an adaptation of levels which goes beyond A2, assuming a cognitive plateau in children who reach this level (Kihlstedt, 2019).

Finally, the exclusive adult perspective on children's English language learning by policy-makers, teachers and also researchers in this field has recently raised some concerns (Pinter, 2011, 2018). In other words, the ethical aspects involved in research and testing with children have been put forward. One could wonder, indeed, whether children's traditional passive role in language education research is justified given the great number of international treaties and agreements, such as the United Nations Convention on the Rights of the Child (The United Nations, 1989), which safeguard this population's rights, ensuring their participation in governance and, hence, providing them with the opportunities to "exercise their agency *in learning"* [italized in the original] (Pinter & Kuchah, 2021). Pinter and Kuchah argue, therefore, that engaging children in research is a guarantee for their right to participation and promotion of social justice, and in turn, their views can enrich both language teaching policies and practice.

1.3.2 The case of Spain and the Basque Autonomous Community

Since the turn of the century, English has been by and large the most studied foreign language in the Spanish education system. In fact, more than 90% of learners in primary, lower- and upper-secondary levels studied it in the period between 2005-2017, according to the latest Eurydice report (European Commission/EACEA/Eurydice, 2017, p. 77). Nonetheless, during the first years of the new millennium the English language national assessments administered to primary and secondary school learners in Spain showed a bleak landscape, as these learners failed to attain the expected L2 levels (A1 by the end of the former and B2 by the end of latter) (Ministerio de Educación, Cultura y Deporte, 2002, 2004). Most of the learners concentrated in the intermediate levels and especially struggled in the areas of oral and written communication. It was in this moment that the Spanish government, following the international trend of EFL practices, decided to set a plan to reverse this trend. On the one hand, the Organic Law of Education (LOE) (Organic Law, 2/2006, of 3rd May) included the obligation to lower the starting age in English learning from 8 to 6 (that is, to the first year of primary education). On the other hand, from 2004 onwards many autonomous communities offered the possibility of implementing Spanish-English bilingual programs (or trilingual in those communities where there is a co-official regional language) (García Bermejo, 2021).

As a result, Spain soon became one of the most experienced countries in Europe in the promotion of early English learning, which was encompassed with the development of CLIL methodologies. As mentioned above, these methodologies involve increasing instruction hours by delivering other subjects in the L2, and hence, they emphasize content and meaning over grammar and form. However, research questioned the benefits of this approach, because there were mixed findings in different linguistic competences, as previously indicated. In fact, scholars suggest that for such rapid implementations to succeed, these should always be accompanied with extensive teacher training programs and the design of appropriate material which supports their instruction (Bland, 2019; García Mayo, 2012).

In any case, the current education curriculum in force LOMCE (Organic Law 8/2013, of 9th December) as well as its successor LOMLOE (Organic Law 3/2020, of 29th December) follow the same trend of maintaining an early start in English while promoting a communicative methodology in the classroom. As most competences in education have been transferred to each autonomous region, these entities are the ones in charge of establishing a starting age for the first foreign language. In four autonomous communities the start is at three years (Cantabria, Castilla-La Mancha, Castilla y León and Comunidad Valenciana). In the case of the Basque Autonomous Community (BAC), whose education scenario will be further analyzed, the curriculum for pre-primary education Heziberri 2020 (Decree 237/2015, of 22nd December, art. 9) gives the choice to each school to set the start date as early as three years or postpone it until six years old, the obligatory starting age (Decree 236/2015, of 22nd December, art. 10).

Regarding the teaching approach fostered in the primary school curriculum, the preamble of LOMLOE (Organic Law 3/2020, of 29th December, p.34) the goal of achieving a basic communicative competence in the first foreign language that enables learners to convey and understand simple messages and manage everyday situations. Heziberri 2020 sets seven main objectives (p. 133), which go from conveying and comprehending oral and written discourses which are related to the school environment or the learners' own interests, to reflecting on the foreign language system. Until LOMLOE defines the new set of contents that are expected to be covered in English for primary education, LOMCE promotes the use of communicative situations circumscribed to children's own life experiences so that it can encourage their motivation to learn the language. Furthermore, it suggests the use of games, especially in the first years, as well as conjoint activities in order to promote the socializing role of the language. Finally, the curriculum indicates that, as children grow older, they can be provided with more theoretical explanations, shifting from a semantic to a syntactic processing. Yet, a recent study (Gris Roca, 2017), which analyzed 100 grammar activities of 10 EFL textbooks in Spain, showed that linguistic content continued to be presented in an implicit way across all six years of primary education.

In the case of Heziberri, which is actually an adaptation of the Spanish law to the specificities of the BAC, the Basque curriculum emphasizes an integrated teaching of the two official languages (Basque and Spanish) and also the foreign language. However, examples of coordinated didactic sequences are still far from being a reality (Apraiz Jaio et al., 2012) and it is most frequent that school policies follow "a strict separation of languages and focus on one language at a time" (Leonet et al., 2019, p. 1).

As far as language policy in the BAC is concerned, there are three linguistic models (Etxeberria & Etxeberria, 2015):

- 'A' model: Spanish is the main language of instruction, Basque is learnt as a second language and English as a FL.
- 'B' model: subjects are delivered in Spanish and Basque (50% of time each). Basque is taught as a second language and English as a FL.
- 'D' model: Basque is the main language of instruction and Spanish is taught as a regular subject, but not as a second language. English is learnt as a FL.

In the latest report from the Basque Government Statistics Foundation (EUSTAT), for the 2020-2021 period (Euskal Estatistika Erakundea, 2021), the results showed a clear inclination towards the D model, as taking all non-university education stages together, more than 68% of learners are enrolled in it, and the percentage raises to 76% in the specific case of Primary Education. In fact, the preference for the D model has been continuous in time. Regarding English, the Basque Government Education Department first introduced it as a language of instruction in 1996 for a pilot program and since then the so-called trilingual CLIL programs (usually a D model with up to three subjects delivered in English) have become increasingly popular (Cenoz, 2012). The Heziberri 2020 decree raised the hours of English instruction in primary education from 2.5 to a minimum of 3. In the light of these figures, we can clearly define the current Basque education system as multilingual.

In summary, as we have explored thus far, the popularization of child EFL teaching and learning all over the world has created countless examples where YLs are acquiring English as their third or even further additional language, hence making multilingualism the norm. Against this backdrop, researchers and instructors should find ways in which English can fulfil some of the key goals set by public institutions (European Commission/EACEA/Eurydice, 2017), such as reflecting on languages, boosting children's interest for *otherness* and, eventually, developing their cognitive and affective aspects, including embracing independent learning and YLs' personality development (Rixon, 2018). Likewise, we should also aim at providing responses to grammatical/linguistic objectives.

In fact, as we have reviewed above, given the specific characteristics of this population when it comes to acquiring an L2 in instructional settings, it is high time that in the field of SLA we started testing and developing pedagogical proposals which are better suited to children's capabilities, and not just overgeneralize from research carried out in totally different contexts. In what follows, we will review the literature of Task-Based Language Teaching (TBLT), and more concretely, collaborative writing (CW) tasks, and we will try to justify why such pedagogical proposals, already proved successful with adult learners, could also be implemented with YLs.

CHAPTER 2 Task-based Language Teaching and Learning

One of the most popularized approaches to language teaching and learning over the last 40 years has been that of Task-based Language Teaching (TBLT). The roots of this approach can be found as far back as the 1970’s, where there was a substantial methodological change in the field of SLA. As Bygate (2015) explains, up to that time, scholars in the field of SLA emphasized the need to capture a broad picture of L2 grammar, for instance, by attempting to determine morpheme sequences or the grasp of particular domains (such as tense, aspect or mood). The traditional methodological approach to obtaining data from the learner’s interlanguage involved analyzing “language in action, rather than studying it through discrete item test” (p. xvii), that is, using large but decontextualized amounts of learner production as a source. Conversely, the opposed trend which emerged at that time argued for the need to use (micro-) contexts in order to analyze how learners make use of meaningful interaction, in particular, while performing specific activities, that is, tasks.

Yet, not only are tasks interesting for SLA research purposes, but they also have a pedagogical relevance. In this respect, TBLT aligns with the principles of Communicative Language Teaching (CLT) (Breen & Candlin, 1980), an approach that originated approximately at the same time as when the aforementioned methodological change took place in SLA (that is around the late 70s and beginning of 80s) (R. Ellis et al., 2020). While initially CLT only constituted a notional-functional mask for a structural syllabus, where grammar was still at the core (i.e. the weak version of CLT), it progressively became more communicative and encouraged assessment in terms of a communicative outcome and with less focus on grammatical accuracy. This later development of CLT (i.e. the strong version), is where TBLT derives from (R. Ellis et al., 2020, p. 4).

In this vein, a task-based syllabus highlights the process of interactive meaning-making, rather than a language learning outcome or product. Hence, according to Bygate (2015), reworking existing knowledge and engaging with the task are two especially important processes in a task-based syllabus, related to Barnes’ (1976) concept of “exploratory language” and Bruner’s (1973) idea of “hypothetical mode of teaching". From this point of view, learners are given the opportunity to formulate their communication by themselves and test their interpretation of the task, while the teacher's role is that of understanding those different interpretations. Bygate considers that TBLT even nowadays challenges the prevailing precepts in education and language learning, and consequently, it still needs to be considered as an innovation.

To illustrate what a task is, a jigsaw might be a typical example, as it has been used extensively in both research (Swain & Lapkin, 2001) and pedagogy (Willis & Willis, 2011). In a jigsaw, the information is split between two learners or groups of learners, so that each of them has a quite different piece of information. To complete the task, they have to put the information together, asking and answering questions about it. This is a classic oral and/or written task involving two-way communication with a clear meaning-oriented goal which is to form a coherent whole text in collaboration.

Just as many other pedagogical and research frameworks, TBLT has been object of criticism and has given rise to different interpretations even among scholar working within the same framework (R. Ellis, 2009; Long, 2016). Many of the (still ongoing) debates have been centered around issues pertaining either to task design or task implementation (i.e. methodology). Most importantly, the argument continues as to what extent grammar should be present in a task-based syllabus concerns both stages. Regarding design, researchers and instructors have to decide whether tasks should be unfocused (with the only goal of stimulating communicative language use) or focused (with a predetermined linguistic target). As for implementation, they should decide if a task will be a one-off activity or part of sequence.

For the current doctoral dissertation, we will follow R. Ellis' definition of what constitutes a task-based approach. R. Ellis (2009) contends that a 'task', the primary unit in TBLT, should meet the following requirements: (i) its primary focus should be on 'meaning' and, therefore, it should require learners to process language semantically and pragmatically, (ii) there should be some kind of information or content 'gap', (iii) learners should resort to their own resources in order to carry it out, and (iv) there is an outcome which exceeds the mere use of language (R. Ellis, 2009, p. 223).

As far as the aforementioned decisions are concerned, R. Ellis et al. (2020) consider that, although initial proposals of TBLT advocated for the use of unfocused task to step away from any resemblance with the structural syllabus, later refinements of this approach (Loschky & Bley-Vroman, 1993) include the possibility of orienting learners towards a particular linguistic feature. Moreover, employing focused tasks does not "result in a return to a structural approach if there is no attempt to teach the target structure directly, only to create a communicative context for its use" (R. Ellis et al., 2020, p. 223). In this respect, these authors also deem the use of focus on form (FonF) justified. FonF is a term coined by Long (1988), who described it as a teaching strategy to sporadically divert learners' attention to linguistic elements while the overarching focus is still on meaning and communication. More recent definitions of FonF describe it as "a set of techniques deployed in a communicative context by the teacher and/or the learners to draw attention implicitly or explicitly and often briefly to linguistic forms that are problematic for the learners" (R. Ellis, 2016, p. 411). Furthermore, FonF can take place pre-emptively (i.e. anticipating a linguistic problem in a specific task) or reactively (in response to a problem that naturally arises during the completion of a task). In fact, there is a variety of strategies to include FonF in TBLT, and the moment of the task cycle when is best to incorporate such component (before, during or after the task) is still an ongoing research question in the field.

As we will see, the use of meaning-making communicative tasks as well as one of their main components, FonF strategies, are grounded on theory and have been thoroughly researched from different scientific perspectives. R. Ellis et al. (2020) acknowledge, indeed, that TBLT should be informed by different strands of SLA scholarship. Although the authors mention up to five different perspectives (cognitive-interactionist, psycholinguistic, sociocultural, psychological and educational) as being influential for TBLT, we will now concentrate on the interactionist, sociocultural and psychological strands, as they provide some of the most relevant constructs which constitute the theoretical underpinnings of the present dissertation.

2.1 The interactionist perspective

Long's formulation of the Interaction Hypothesis (IH) (1996) still guides much of the research generated nowadays in SLA, and TBLT is no exception, although research questions have certainly widened, their scope and more variables (such as technology mediated interaction) are currently considered. According to the IH, negotiation for meaning, especially that involving interactional adjustments by a native speaker or a more proficient interlocutor, facilitates acquisition, "because it connects input, internal learner capacities, particularly selective attention, and output in productive ways" (Long, 1996, p. 451-452). This perspective was revisited by Gass and Mackey (2006), who withdrew the original requirement that effective interaction should necessarily occur with a more competent interlocutor.

One of the main concerns of the interactionist approach is related to the sort of learning fostered by tasks which provide opportunities for input and interaction. This is, in fact, when Schmidt's (1990) Noticing Hypothesis comes into play. This hypothesis predicts that in order to learn any aspect of the L2, learners need to notice the new linguistic data in the input. Therefore, Schmidt contends that some sort of attention is necessary in order for learning to take place. How much attention is necessary is another question SLA researchers need to solve. Initially, Schmidt suggested that there are two different levels of attention, a shallower one, *"noticing"*, and a deeper one, *"understanding"*. Although Schmidt considered in the first version of his hypothesis that learning without noticing was not possible, he later (Schmidt, 2010) conceded that this was possible at least to a certain extent, by means of what he labelled "detection".

The concept of 'noticing' is key to understand the difference between implicit and explicit learning. The former can occur when learners learn without conscious attention to linguistic forms in the input (i.e. without noticing), and results in implicit knowledge, deemed necessary for fluent communicative language use. However, there is another route to achieve implicit knowledge, and that is by means of explicit learning "which serves as activator of noticing and, in this way, facilitates the development of implicit knowledge" (R. Ellis et al., 2020, p. 31). This activation is in no way automatic and may actually never take place, as it contingent on external factors (such as FonF techniques employed to attract learners' attention or the opportunities for producing output) and on internal factors (for instance, working memory) (N. Ellis, 2005; R. Ellis, 1994).

Finally, R. Ellis et al. (2020) include another type of learning, known as incidental learning, which may take place under implicit or explicit instruction. As an example, they cite the study by Loewen et al. (2009), which explored if the 3rd *-s* (discussed below) could be incidentally acquired when learners were provided with instruction on a different feature (i.e. indefinite articles). In fact, R. Ellis et al. (2020) explain that this is the kind of learning that occurs when learners are engaged in a communicative task where meaning is central and they sporadically divert their attention to form. Within the interactionist perspective of TBLT, L2 implicit knowledge is acquired by means of incidental (and not intentional) learning, that is, when learners are assisted to attend to form while they are communicating (what is also known as FonF).

A second concern for scholars working from the interactionist perspective is the role of output in L2 learning. In fact, one of the critiques to TBLT is the impoverished quality of interactions that emerge between learners when they are performing tasks, especially when learners are beginners (Seedhouse, 1999). To illustrate what one could understand as poor (or little developed) interaction we show two examples of the current dissertation data, where two child learners (ages 11-12) are interacting in a collaborative writing task:

(1) [ABE007 & ABE011 - Day 1]

*CHI A: but eh… in the supermarket.
*CHI B: on the supermarket.
*CHI A: on *sí* [yes]

(2) [ABE012 & ABE022 - Day 1]

*CHI B: *en vez de comprar un* (instead of buying an) apple they…
*CHI A: apple?
*CHI B: no, they… *en vez de comprar un* (instead of buying an) apple they…
*CHI A: they eh… *¿cómo se dice comprar?* (how do you say "buy"?)
*CHI B: shop? *¿o cómo era?* (or how was it?)
*CHI A: shopping.
*CHI B: shopping don't the apple.
*CHI A: shop the apple.
*CHI B: apple.

In (1) we see that the two learners encounter a problem in the use of a preposition, but given that they lack the necessary linguistic resources in the L2, they simply exchange two options aloud ('in' and 'on') until they reach a decision, which is incorrect. In (2), although the interaction is longer, the learners are not able to come up with a successful equivalent for their desired construction in the L2, and once again we perceive an exchange of alternatives (‘shop’, ‘shopping’), without any use of metalanguage.

However, R. Ellis (2009) would contend that even such cases can constitute occasions where learners force themselves to use their limited resources and develop strategic competences. In fact, *pushed output* can facilitate noticing, and several authors have stressed the positive influence of production on language acquisition. Swain (1985, 1995) identified three principal advantages in output: (i) it promotes noticing ‘the gap’ between what learners want to say and what they can say in the target language, (ii) it involves hypothesis formulating and testing, as learners try out new language forms and, as a consequence, can modify their interlanguage, and (iii) it raises metalinguistic awareness by making learners use language to reflect on their own language use. On a subsequent revision, Skehan (1998) also included as another role of output the fact that it serves to develop automaticity of existing L2 knowledge and considered its input transforming role, highlighting that learners’ efforts in production provide certain clues to modify the input. Last but not least, R. Ellis et al. (2020) also consider the possibility of “auto-input” that output offers to learners when they are interacting.

The language and discussions learners produce while carrying out tasks have been the object of the so-called interaction studies. This interaction can occur both between the teacher and the learner, but also between two or more learners. In the first stages, interaction studies concentrated on scrutinizing discourse strategies and, more specifically, classifying discourse signals (or indicators) which reflect learners’ meaning making efforts. Apart from the interactional moves aimed to check communication (repetitions, paraphrasing, etc.), some researchers distinguished another kind of discourse indicators which deal exclusively with linguistic elements, and hence, belong to *negotiation of form* (Lyster, 2001).

Negotiation of form is tightly linked to the concept of FonF, introduced above. Depending on the nature of the task (whether it is focused or not), the linguistic discussions that arise will be more or less predictable. Thus, focused tasks, which are designed for pre-emptive focus on form, have the advantage of allowing for pre-testing and post-testing of the targeted forms. Yet, even in such tasks, research has shown that learners tend to go well beyond the targeted features. As Bygate (2015, p. xix) "a task is not a mechanism which operates unfailingly on learners", and many other variables should be taken into consideration.

Although researchers in the interactionist framework have traditionally concentrated on the study of negotiation of meaning and reactive focus on form (resulting from corrective feedback), spontaneous linguistic problems discussed by learners have also been analyzed. The way to operationalize those instances has usually been language-related episodes (LREs), that is, "any part of dialogue where the students talk about the language they are producing, question their language use, or correct themselves or others" (Swain & Lapkin, 1998, p. 326). In a recent meta-analysis (Plonsky & Kim, 2016), it was found that a quarter of all interactional features appearing in TBLT publications were LREs. In fact, this unit of analysis, far from being exclusive of the interactionist approach, is also shared by researchers working from the sociocultural perspective, as it draws on the view that, as will be further explored below, collaborative dialogue is not only a means for facilitating learning, but rather that it constitutes learning in progress.

2.2 The sociocultural perspective

The social dimension of TBLT did not take off until the 90s (R. Ellis et al., 2020), encouraged by the work of Swain and her colleagues (Swain, 1985, *et passim*). Tasks, regarded until then as tools which triggered cognitive processes that facilitated acquisition, began to be considered as the actual site where learning took place. The theory that was most influential for SLA scholars was the Sociocultural Theory (SCT), proposed by different authors from the soviet school, such as Vygotsky (1978) or Leontiev (1981), that is, before the "social turn" in TBLT emerged. Although SCT scholars referred to learning in general, many of their theoretical proposals and constructs are used to explain sociocultural SLA.

From this perspective, the social use of language is paramount, as it constitutes the source and context for higher-order abilities, or in terms of knowledge, the means to acquire scientific concepts. These concepts differ from everyday concepts in that an individual cannot grasp them by him/herself, but instead, he or she needs mediation.

Mediation is, indeed, a central construct in SCT: learners need social artefacts to mediate certain activities, such as learning higher-order abilities. The social interaction that occurs between an expert and a novice (usually teacher-learner dialogue), but also among learners themselves, establishes the grounds for that mediation to happen. In SCT terms, it helps to co-construct a Zone of Proximal Development (ZPD), a metaphorical space between the learner's level of current ability to solve a particular problem and the potential one, which can be achieved with the careful assistance of someone else. Moreover, mediation can also occur individually, through private speech, that is, by means of self-regulation.

The joint interactional behavior through which learning is regulated and ZPD co-constructed has also been called 'scaffolding' (Wood et al., 1976). Scaffolding refers to "the assistance to perform a skill or a linguistic feature one cannot perform by him/herself" (R. Ellis et al., 2020, p. 108). Although initially it was assumed that such process could only take place between the "novice" and the "expert" in a unidirectional way from the latter, understood as the teacher or instructor, to the former, this definition was later reviewed (Ohta, 2001), in order to include those contexts where scaffolding takes place between true peers, with the same or a different level of command over the knowledge they are discussing.

As far as scaffolding about language is concerned, Swain (2006) coined the term "languaging", and defined it as a "a dynamic, never-ending process for using language to make meaning" (p. 96). In fact, she highlighted both the meaning-making function and the knowledge and experience shaping role of language. A large body of research has been devoted to analyzing the quantity and characteristics of languaging originated during task performance. The most usual unit to operationalize these discussions has been the aforementioned LRE (Swain & Lapkin, 1998). LREs have been classified according to their linguistic focus (grammar, spelling, lexis, pronunciation, etc.) and their resolution (correctly or incorrectly resolved, or alternatively, left unresolved). The following example found in the study by Swain and Lapkin (1998, p. 332) with a pair of adolescent learners of French provides an example of an LRE generated during a collaborative writing task (a jigsaw, explained below), which they completed after a mini-lesson about French reflexive verbs:

(3)

Rick: *Un bras…* wait… *mécanique…* sort? (An arm… wait… a mechanical [arm] comes out?)
Kim: *Sort*, yeah (Comes out, yeah.)
Rick: *Se sort?* (Comes out?) [incorrect reflexive form]
Kim: *No, sort* (No, comes out.) [correct nonreflexive form]

In (3), both participants engage in metatalk (languaging) about this specific grammatical feature, as Rick wonders about the linguistic nature of *"sort"* (*'sortir'*, 'to come out') and Kim's assistance is necessary to determine the correct form of the verb (nonreflexive). Swain and Lapkin (as other SCT-oriented studies which have analyzed collaborative dialogue during task completion) employed a pretest and a tailor-made posttest (based on the learners' LREs) to measure how much learners had benefitted from collaboration. They concluded that the number of LREs and the posttest scores were significantly correlated, suggesting an influence of languaging on L2 learning.

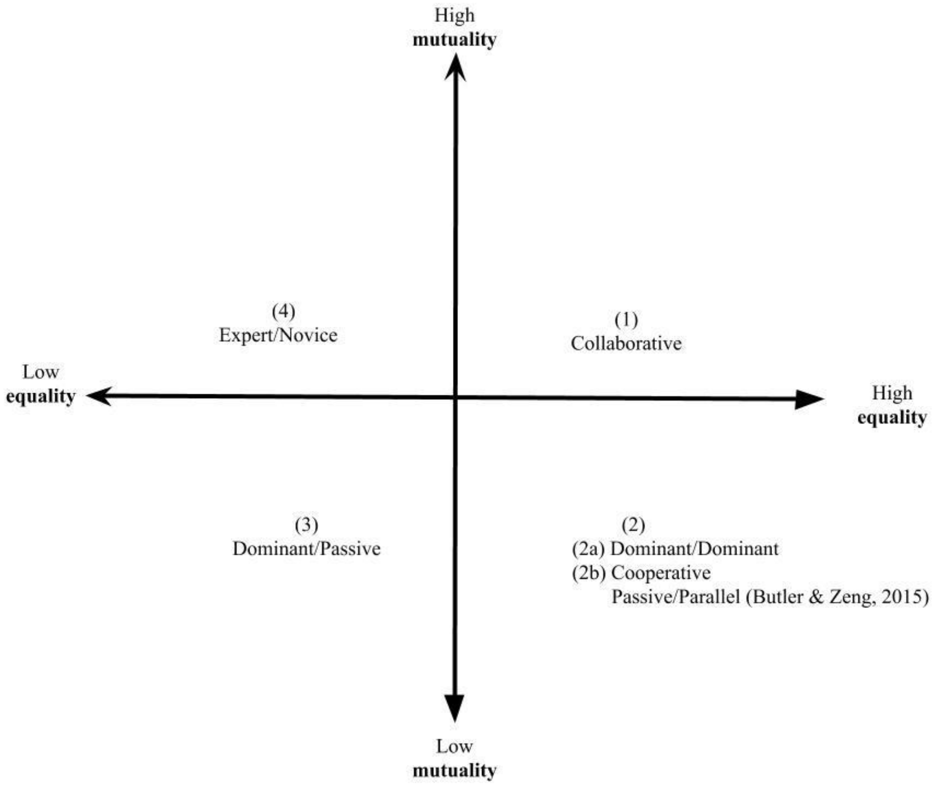
At this point, it is worth noting, as pointed out by R. Ellis et al. (2020), that LREs, as well as the proper definition of scaffolding and mediation entail rather explicit attention to forms and intentional language learning. This preference marks a stark difference compared to the cognitive-interactionist perspective about the nature of learning that TBLT should be fostering. Contrary to the idea of promoting implicit learning (Long, 2007), SCT tenets of TBLT favor explicit teaching and learning with consciousness, and hence, these principles are defended by sociocultural scholars (Lantolf, 2012; Lantolf & Thorne, 2006). In fact, sociocultural SLA refuses any sort of developmental path and contends that any linguistic form can be taught, regardless of its complexity. However, research from this perspective has not yet provided many examples of how the learning originated from interactional dialogue becomes automatized (R. Ellis et al., 2020). In the study by Swain and Lapkin (1998), for instance, pretests and posttests involved gap filling, grammaticality judgments and multiple-choice items, that is, rather constrained contexts of language use. Thus, there remains a need to examine if that co-constructed knowledge is mirrored in spontaneous speech (oral or written).

The impossibility to predict the outcome of a particular task, however, is well rooted in sociocultural TBLT. As R. Ellis et al. (2020) explain, this is the main reason why

SCT provides very little guidance with regards to how a task-based syllabus should be organized and tasks designed. Storch (2017), for example, argues that tasks should be challenging (i.e. they should avoid delving into everyday concepts and aim at higher order ones) and contain effective support (including the means to co-create a ZPD). Yet, sociocultural TBLT claims that learners bring a range of “goals, actions, cultural background, and beliefs (i.e. their agency) into tasks, and thus transform them” (Donato, 2000, p. 44), hence, making these tasks unpredictable to a large extent. Unpredictability is also recognized by cognitive-interactionist scholars such as Bygate (2015), as we saw in his quote above, acknowledging an inherent flexibility of tasks, regardless of how fine-grained their design is.

Studies that have analyzed languaging and LREs in particular have also attested high variability among learners. SCT scholars (Storch, 2002; Swain et al., 2009) have claimed that learners’ ability to produce more and higher quality LREs is dependent on the patterns of interaction during collaborative work. Storch (2002), drawing on previous research by Damon and Phelps (1989), was the first scholar that investigated this collaborative dialogue in SLA in relation to two concepts: equality and mutuality. The former refers to the learners’ “level of contribution and control over the task”, whereas the latter constitutes the learners’ “level of engagement with each other’s contribution” (Storch, 2016, p. 393). The different combinations of equality and mutuality are illustrated on a matrix quadrant similar to the one in Figure 1:

Figure 1 Patterns of dyadic interaction (adapted from Butler & Zeng, 2015; Storch, 2002, 2009a, 2016)



In quadrant (1) above, both members of the pair engage in collaborative dialogue (i.e. scaffolding each other) to the same degree and engage with each other's proposals. In quadrant (2), although there is a contribution to the task from both sides, they are too concerned with their own suggestions and do not pay attention to their partner's, hence generating conflicts which are far from being constructive discussions. The low mutuality/high equality combination, however, can also result in a cooperative pattern, where both learners make contributions, but do not engage in each other's, as if they were only responsible for their own part (Storch, 2013). The cooperative pattern has also been named passive/parallel in research with YLs. In fact, Butler and Zeng (2015), who studied the interaction of 48 Chinese EFL YLs (ages 9-12) in two oral decision-making tasks, argue that the label "cooperative" can lead to a confusion as it misrepresents the lack of or practically non-existent engagement in the task interaction.

Next, in quadrant (3), one member of the dyad takes the task upon him/herself, and leaves very little room for the partner's contribution. Finally, quadrant (4) represent the typical interaction contemplated by SCT principles, that is, a teacher/learner relationship; despite one learner taking the leading role, there is enough assistance and support provided so that a ZPD is co-constructed. Interestingly, however, it is pattern (1) that has been found most effective in terms of L2 learning (Storch, 2017).

Furthermore, collaborative patterns have been claimed to be dependent on the learners' proficiency. Some researchers have found that they are more frequent among proficiency matched dyads (low-low or high-high) than in different proficiency ones (high-low) (Storch & Aldosari, 2013). Nevertheless, other studies show little influence of proficiency on the nature of learners' languaging episodes (Watanabe & Swain, 2007). What is quite clear for sociocultural scholars is that proficiency is not the only reason that accounts for dyadic patterns. In fact, as we will review in the next section, devoted to the psychological perspective, learners' goals and attitudes towards a particular task can impact on task dynamics and outcomes.

2.3 The psychological perspective

As we have already seen in the review of the cognitive-interactionist and the sociocultural paradigms, tasks and the variables related to their design and implementation (e.g. form-focused vs meaning focused, involving an information gap or not, oral or written etc.) are reinterpreted by learners. In fact, learners may (and often do) approach the teacher's or researcher's goals very differently. This divergence is very much related to the cognitivist distinction between task-as-workplan and task-as-process, in other words, the difference between task design and implementation (Breen, 1989). In addition, the sociocultural strand refers to the implication of emotions in any kind of learning. In fact, the construction within ZPDs and mediation in learning generally take place along with learners' positive emotions (R. Ellis et al., 2020).

The learner-related factors that come into play in the performance of tasks have commonly been labelled "individual differences" (IDs) and have been explored with the help of theories and constructs from the field of psychology. More specifically, these IDs can be defined as the "mental experiences, processes, thoughts, feelings, motives and behavior of individuals involved in language learning" (S. Mercer et al., 2012, p. 2). IDs can be classified into two broad groups: cognitive and affective. The former consists of variables such as language aptitude or working memory (Granena et al., 2016), whereas the latter have been mainly explored in terms of motivation and anxiety. For the purpose of this dissertation, we will now concentrate on affective individual variables and analyze their relevance for TBLT. Learners' emotional dimension towards tasks, in fact, is now a pivotal theme within this framework (Lambert, 2017).

The study of affective individual variables in SLA began in the 90s, along with the popularization of CLT, which assumes that the learner plays a central role in the process of acquiring an L2. Hence, researchers started to be interested in the learner's needs, expectations, goals, motivation and beliefs (S. Mercer et al., 2012). One of the most researched constructs has been that of L2 motivation. In fact, it is considered a topic inherent to L2 research, as success in the second or foreign language acquisition process has been demonstrated to depend largely on it (on the contrary, no one considers how motivated individuals are when acquiring their L1). According to R. Ellis (2015), motivation is a complex construct consisting of three different components: (i) the reason of learning an L2, (ii) the effort an individual invests in the learning process and how it is influenced by the immediate context, and (iii) the impact of the evaluation of the outcome and progress of learning on subsequent behavior. These three components are grouped into two categories: the first component corresponds to the "macro" level, whereas the second and third components fall under the "micro" category.

Drawing on theories from educational psychology, the earliest studies on L2 motivation carried out in the bilingual immersion programs in Canada (Gardner & Lambert, 1972) found solid arguments to consider L2 motivation an independent domain. In particular they referred to the strong relationship between language and identity with a particular community, involving both psychological and social dimensions that are not present when acquiring or learning other kind of knowledge. Nonetheless, the field has evolved since then, and it is now acknowledged that L2 motivation studies should converge with general motivation studies (Ushioda, 2012), as well as make a clearer emphasis on the dynamic relationship between the self and the context. This shift has been considered a transition from the social-psychological to a process-oriented perspective (Csizér, 2017).

One of most influential work in the field of L2 motivation has been that carried out by Dörnyei and his associates (Dörnyei, 2002, 2005; Dörnyei & Otto, 1998; Dörnyei &

Ushioda, 2009). Dörnyei defends a dynamic model of motivation, that is, assuming that, influenced by various inner and outer factors, this construct is subject to temporal fluctuations. According to the L2 motivation Self System theory (Dörnyei, 2005), the amount of effort invested into language learning will mainly depend on three variables: the ideal L2 self (the self-concept learners have as successful users of the L2), the ought-to L2 self (outside pressures throughout the learning process acknowledged by learners), and, finally, the L2 learning experience (including those factors affecting classroom processes, such as the school, course, instruction, class, target language, etc.). R. Ellis et al. (2020) suggest that, when referring to task motivation, the last component, the L2 learning experience, is most relevant. Indeed, learners display certain attitudes towards tasks, they have their own perceptions about task difficulty and complexity, and their views are subject to change when working along with other participants.

Several pieces of empirical research, for instance, established a relationship between the L2 learning experience motivation and the performance in interactive tasks. Dörnyei (2002) studied forty-four Hungarian secondary school EFL learners' general motivation and task-related motivation, and examined them in relation to the number of words and turns produced during an oral dyadic task. He found that all types of motivation (except integrative, which is part of the ideal L2 self) significantly correlated with the number of turns. Yet, when subjects were classified as high-task attitude or low-task attitude learners, he found that task-specific motives were only predictive for the high-task attitude, whilst course-related motives explained more of low-task attitude learner behavior. In a way, the latter compensated their lack of interest in the task with their high interest in the course. Furthermore, it was also noteworthy that low-task attitude learners were influenced by their interlocutors' motivation. Using the same sample, Kormos and Dörnyei (2004) showed that task attitudes were correlated with the number of oral turns, but not with the quality of those turns (i.e. complexity and accuracy).

Research in instructed second language acquisition (ISLA) has traditionally placed a greater focus on learner-related motivation factors (analyzing their disposition and preconceived attitudes towards the L2 learning process) than on the impact that the actual instruction (i.e. L2 learning experience) can have on the learner (Csizér, 2017).

However, the role of teachers in creating tasks which mediate learners' motivational level is now widely assumed. Likewise, it is now more common to find studies investigating how certain task design variables might be related to that context-specific motivation, which has been operationalized in terms of engagement (Lambert, 2017).

Dörnyei (2019) himself acknowledges that not enough attention had been paid in his L2 self-system theory to the third component, the L2 learning experience, and argues that the concept of engagement is the one that can best capture it. Although still lacking theoretical consolidation and posing problems regarding measurement (Lambert, 2017), Dörnyei provides some clues as for what we should understand by engagement. In the field of educational psychology, 'engagement' refers to active participation and involvement in certain behaviors. This is regarded as key in SLA, since in order to achieve the goal of automatization of L2 knowledge, learners' meaningful participation is necessary. In this vein, both CLT and TBLT consider that active student involvement comes about through "learning-by-doing".

L2 learning experience is, hence, revisited as the "perceived quality of the learners' engagement with various aspects of the learning process" (Dörnyei, 2019, p. 25). Engagement consists of both an external and an internal dimension, hence, is observable through the learners' behavior, but it also has certain internal aspects more related to their cognitive and emotional participation. Consequently, different accounts of engagement in SLA tend to consider it a multifaceted construct, consisting of a behavioral, a cognitive, an affective and a social dimension (Philp & Duchesne, 2016; Svalberg, 2018). Yet, Dörnyei (2019) posits, in accordance to other educational psychologists (Skinner et al., 2008), that it is the behavioral aspect we should really be looking into. In fact, that is what really distinguishes engagement from motivation, as the latter indicates a *potential* for certain learning behaviors, whereas the former constitutes the actual "behavioral outworkings of various motivational sources" (Henry & Thorsen, 2018, p. 3).

Having specified the notion of engagement, Dörnyei (2019) goes on to list several facets L2 learners can engage with, and thus, that research should aim to capture or measure. These include the school context, the syllabus and the teaching materials, one's peers, the teacher and, more relevantly for TBLT, the learning tasks. In the special issue devoted to learners' engagement with tasks (Lambert, 2017), two task design variables were identified as engagement boosters. Firstly, the power of decision learners have over the task content: the more control, the higher engagement. Secondly, drawing learners' attention to meaning in detriment of form. However, Lambert acknowledges that more research is needed including different learning contexts and task types.

To summarize this section where we have explored the psychological perspective on TBLT, the focus has been on how affective IDs, and more specifically, motivation and engagement, can have an impact on task performance and learning outcomes. Clearly, the study of affective variables in relation to tasks is currently a burgeoning field and we also intend to shed some light on it in the present dissertation.

CHAPTER 3 Task-based Language Teaching and Learning with children

3.1 Theoretical and pedagogical factors surrounding TBLT and YLs

In the previous review of TBLT and its main theoretical paradigms, several pieces of empirical research have been mentioned. However, these findings were based on adult or adolescent L2 learners. We now aim to offer a summary of TBLT research to date with YLs. Indeed, if we consider the main principles underpinning the implementation of tasks in the L2 classroom, we could argue that they favor the child population's characteristics. Firstly, tasks encourage implicit learning processes, which are considered the primary learning mechanism in children in contrast to adults, who rely more on explicit language learning (DeKeyser, 2000, 2003; DeKeyser et al., 2010; Muñoz, 2006; but see Lichtman, 2016; Roehr-Brackin & Tellier, 2019 for the impact of the type of instruction in promoting explicit L2 learning in children).

Secondly, tasks tend to foster interaction and collaboration among language learners, and this can certainly create favorable learning conditions for children, who are inherently inclined to communication, as well as play and fun, rather than accuracy (Halliwell, 1992). In primary school years, in fact, children start participating in different types of school talk (N. Mercer, 1995), including disputational talk (characterized by disagreements), cumulative talk (aimed at building positively upon what the interlocutor said) and, finally, exploratory talk, where learners critically engage in the construction of discourse by offering or requesting suggestions and alternatives. Furthermore, there is also room for FonF strategies which draw learners' attention to linguistic features while completing tasks, since during middle childhood metalinguistic thought rapidly increases, and children are more able to bring implicit knowledge into consciousness, analyze unknown words, and make inferences (Pinter, 2011).

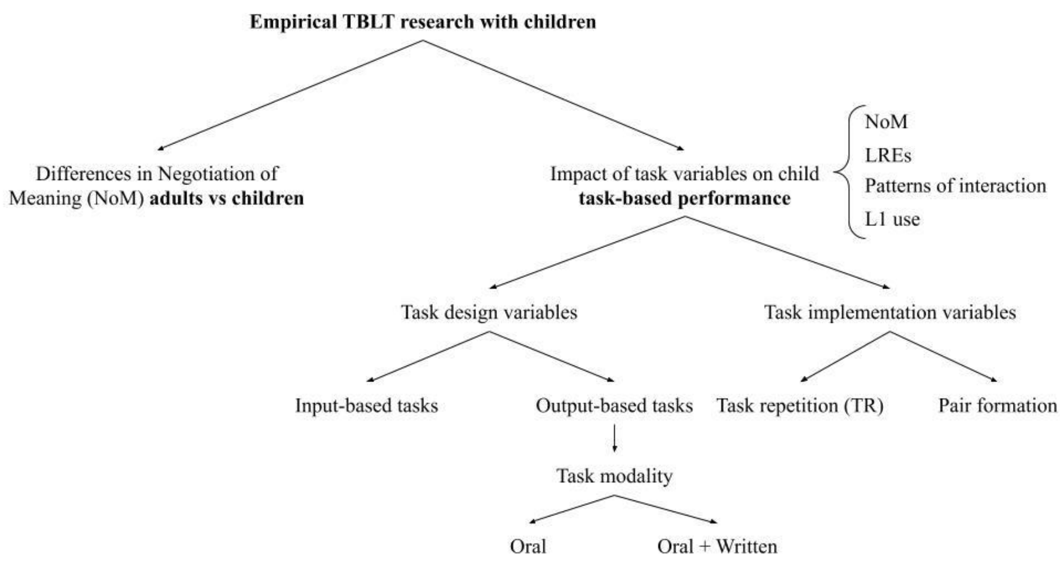
Last but not least, as explained from the psychological perspective, learning an L2 is not just a matter of cognitive abilities, but it also entails emotions and affective factors. In this respect, although tasks can be face-threatening for many adult or adolescent learners, as their performance can be judged by their peers and they may feel that this learning approach interferes to some extent with their own identity, there are reasons to believe that this concern is less relevant in the case of children. YLs have a less marked identity and they are often described as “learning optimists”, that is, they tend to have a higher consideration of their own abilities. At the same time, older children start experiencing lower self-esteem and motivation more typical of adolescents. Interpersonal relationships, such as friendships, also become more complex as learners grow older, and children start being more selective on who they consider their friend. By the ages 11-12, sharing interests, reciprocity and mutuality become more important features of friendship. As Pinter (2011) explains, these are all matters to be taken into consideration when implementing tasks, since a secure and friendly environment helps YLs to perform better.

TBLT-framed pedagogy, although not being originally conceived for YLs, has long made recommendations regarding the implementation of tasks with this population. One of the earliest recommendations can be found in Willis’ (1996) handbook. In fact, she devotes a whole chapter to children and tasks. Her recommendations include starting with input-based tasks (such as listening tasks) to help these learners acquire the language more naturally. The author emphasizes throughout the chapter the need to focus first on lexis and move progressively onto grammar, but no focus on form ability is foreseen in children. Furthermore, she recommends including a pre-task stage as a way to introduce the topic and provide (or pre-teach) some of the language that will appear in the task. Although Willis covers an age range up to 12 years old, the tasks proposed take the form of simplified tasks (e.g. reading a story and afterwards ordering jumbled cards with sentences from the story) or are closer to pure games (such as memory card games or “Simon says”). Implementing a task-based language syllabus with YLs, however, is not without difficulties. In the specific case of young EFL learners, Carless (2002) identified the main factors related to the disadvantages of implementing tasks in this setting. These included cramped classes (which hinder a needs analysis), teachers’ lack of training, the difficulties involved in keeping discipline, an excess of noise, a limited pupil involvement (usually restricted to certain individuals) and also an excessive use of the L1. Despite the fact that all these factors may bear some truth to reality, they have been shown to be somewhat unsubstantiated, as a meta-analysis failed to find differences in effect sizes of the effectiveness of TBLT pedagogy with respect to school institution type or region of the world where it was implemented (Bryfonski & McKay, 2019).

3.2 Empirical TBLT research with YLs

As the main factors related to TBLT and YLs have been considered, we will now review the still scarce body of empirical studies that have examined how child L2 learners fare with different tasks. Indeed, Willis' (1996) and other educational scholars' recommendations usually rely on their own vast experience as language practitioners, but these intuitions need to be supplemented with findings from empirical research, in order to shed some light on "what type of language tasks children at different levels of development can cope with and enjoy" (Pinter, 2007b, p. 134).

Figure 2 Summary of the review of empirical TBLT research with children



As Pinter (2015) explains, there are two major strands in TBLT research with YLs. As summarized in Figure 2, one of them is devoted to showing how different or similar children and adults are when negotiating for meaning in information-gap tasks. The second strand investigates how a wide range of task variables affect children's task performance. In what follows, both strands are explored in detail and the summary of the main findings from empirical studies is provided. Finally, this section ends with some general conclusions drawn from the body of research reviewed.

3.2.1 Comparison of task-based interaction in adults and children

The largest body of research corresponds to the first strand. These studies usually classify interactional moves according to a well-established taxonomy in the field (Long, 1983; Pica & Doughty, 1985), including clarification requests (aimed at *clarifying* the preceding utterance), confirmation checks (when the listener tries to demonstrate that the interlocutor's preceding utterance has been heard and understood), comprehension checks (when the speaker tries to establish whether the interlocutor has understood the previous message) and acknowledgements (when the listener confirms that they have understood the preceding utterances using "I understand", "ok" or similar expressions). Other studies also include corrective feedback indicators, such as metalinguistic clues, repetitions, elicitations and recasts (see Lyster & Ranta, 1997).

One of the earliest examples is the work by Oliver (1998, 2002), in the Australian English as a Second Language (ESL) context, who compared the interactions of L2 English children[1] pairs (n = 32, aged 8 to 13) with the adult L2 interaction reported in previous study by Long (1983). Learners had to complete two oral tasks: the first was a picture description task, while the second was an information-gap task. Yet, the results were provided overall for the two tasks. Oliver observed that the children's strategies differed to a certain extent from those of adult learners. For instance, children used far more self and other-repetitions while comprehension checks were much fewer than in adults. This difference in YLs' negotiation was attributed to their egocentric behavior, as children strive more to convey their own message than to understand their partner's. In fact, such behavior became even more evident when analyzing younger children's (aged 5-7) negotiation, who not only used fewer comprehension checks, but also fewer repetitions (Oliver, 2009).

Another study set in an ESL context (Mackey et al., 2003), studied the differences in the response to the provision of feedback by a group of child dyads in Australia (n = 12, 8-12 years old) and adult dyads in the USA (n = 12, fourth-year college students), both groups having a lower-intermediate proficiency in English. Participants completed two information-gap tasks: a one-way task (where only one member of the dyad had

[1] Although these two studies included data from English L1 children (of the native-speaker and non-native speaker condition) only those from L2 YLs will be reported.

the whole information) and a two-way task (where the information was shared between the two members). As in the aforementioned study, the results of the two tasks were analyzed together. Their findings showed that children generated significantly more modified output than adults. Yet, the general picture resulted in more similarities than differences, as both adults and children were able not only to overcome communication breakdowns but also to provide and make use of feedback.

Regarding EFL contexts, two studies were conducted by Pinter (2006, 2007b) in the Hungarian context. She compared the performance of adult dyads (n = 5, college students) and child dyads (n = 10, 10 years old) with a low proficiency level in English. These learners had to carry out an information-gap referential task (spot the differences) three times. The analysis focused on the task outcomes and strategies. Pinter found that both groups managed to interact and complete the tasks, but there were several differences in their task-solving strategies, the amount of language production and their reliance on the L1. Adults in general showed a larger array of strategies while children produced slightly more turns in their mother tongue.

A more recent study by Lázaro Ibarrola and Azpilicueta (2019) also compared child-child and adult-adult interactions in the Spanish EFL context. The 10 child pairs (mean age 8.5) and 7 adult dyads (mean age 47) had an A1 (elementary) proficiency in English. They were presented with the same two communicative oral tasks: two five-picture story rearrangement cards (where one member was the information holder and the other, the receiver, in charge of ordering the jumbled pictures). The authors looked at the types of negotiation of meaning (NoM) and their rates, as well as the main communicative functions those strategies served. Their findings were in line with those from Oliver (1998, 2002 and 2009), as learners were shown to be using the same set of conversational strategies as ESL children, including conversational adjustments, acknowledgments and repetitions. Despite the fact that the authors did find statistical differences between adult and child pairs, they believed that their dataset provides clear evidence for the commonalities (for example, children and adults generated the same amount of conversational adjustments). Their results converge with previous literature by showing that children make less use of confirmation checks, but conversely that they resort more often to clarification requests. Yet, the authors did not find evidence to support claims of a more egocentric behavior of children with respect to adults, as both groups used comprehension checks to the same extent (i.e. were equally concerned that their interlocutor was understanding what they were saying). Likewise, the authors also found that acknowledgments were the most frequent strategy in both groups and, regarding repetitions, both preferred self-repetitions to other-repetitions. Finally, as far as communicative functions were concerned, their results showed that repairs were scarce in both groups, whilst children and adults were committed to reinforcing their message and preventing misunderstandings. They concluded from these findings that child and adult learners were aware of their partners' communicative needs to the same extent.

3.2.2 The influence of task-related factors on child task-based performance

A second body of research on TBLT and YLs has been devoted to studying the impact of different task design and implementation variables on children's negotiation of meaning and capacity to focus on form. As explained previously, there are a number of decisions involved in task-based instruction and research, some of them are related to task design and others to their implementation.

Regarding task design variables, it should be noted that task types can be classified as input-based or output-based. The former require learners "to process the oral or written information provided and demonstrate their understanding of it (for example by drawing a picture or making a model" (R. Ellis et al., 2020, p. 12), whereas the latter require "the learner to speak or write to achieve a task outcome" (p. 12). Although in TBLT research output tasks have traditionally prevailed (Révész, 2019), it is acknowledged that at beginner levels input-based tasks should be employed, given that learners still lack the necessary linguistic resources to produce the L2 meaningfully.

One of the first studies looking into how manipulating input (via input enhancement) could help young learners' development of a specific grammatical feature (L2 English questions) was carried out by L. White et al. (1991). In the context of English immersion in francophone Canada, a group of children (n = 53) was provided with FFI (explicit instruction 3 h + corrective feedback 2 h) on questions, while the control group (n = 76) followed a different instruction on adverb placement. In the post treatment tests (where learners had to determine whether the question order was grammatical or not), the authors found that input enhancement helped child learners to develop accuracy in L2 English questions.

Another study that considered input tasks with YLs was conducted by Shintani (2012), in the Japanese EFL context (n = 30, 6 years old, beginners). Half of the children carried out three input-based listening-do-do tasks and the rest, the control group, took part in different kinds of lessons (such as learning songs, the alphabet or Total Physical Response). By means of a pretest and two posttests (assessing comprehension and production) the researcher tested their uptake in 36 meaning-bearing lexical items and the plural -*s* (present in 6 items). Furthermore, the teacher targeted the plural marker with reactive corrective feedback in response to the children's interventions when asking about the meaning of vocabulary. The input tasks, which took place over five weeks, were the following: in Task 1 children reacted to the teacher's requirements to choose the picture cards depicting the target items and place them in the correct context (a zoo or a supermarket); in Task 2 learners also had to identify the cards (animals or objects) in relation to the teacher's statements and the non-linguistic goal was to help the animals to find the objects; and, in Task 3, a picture bingo was used with the cards from the two previous tasks. The results show that learners' self-directed and other-directed speech during the tasks was facilitative of language learning, and while the former was closer to self-regulation strategies, the latter involved negotiation of form and FonF. The input-based group showed an advantage over the control group in the learning of vocabulary. However, this advantage was not so clear in the learning of grammar. Although the experimental group achieved greater comprehension gains in the plural -*s*, they did not differ in production gains.

The learning of grammatical features through input-based tasks has also been studied more recently by Kasprowicz and Marsden (2018). Their context was German as a foreign language in the United Kingdom (n = 138, ages 9-11). In particular, they wanted to explore to what extent engaging in form-meaning connection task-essential practice (i.e. spotting the target feature and answering questions whose answers depended on the correct interpretation of the meaning of the target linguistic features) was more or less beneficial than simple task-essential noticing practice (i.e. simply spotting the target features and answering non-grammar related questions, for example, about the meaning of nouns and verbs). It must be mentioned that during the five treatment sessions, prior to the listening and reading input-based task, learners also received a short explicit information presentation about the target forms (German masculine definite article case marking). Using six measures in a pretest, posttest and delayed posttest the authors demonstrated that noticing forms could be just as beneficial as establishing form-meaning connections with regards to learning uptake (at least at an explicit level, as the authors could not guarantee they tapped into learners' implicit knowledge).

Turning now to output-based tasks, a recent study by García Mayo and Imaz Agirre (2019) assessed which task modality, speaking or speaking + writing, could promote more FonF and NoM among children dyads. Thirty-two pairs of children aged 11-12 with an elementary level in English completed two different decision-making tasks with one-week gap in between. In the first one, children had to reorder some pictures to form a coherent story. The second task, conversely, can be classified as a decision-making task. It consisted in collaboratively analyzing an illustration containing the picture of a messy laboratory and some suspects, and afterwards, making a guess in writing about who the culprit might be. The learners' instances of attention to form were operationalized in the form of LREs and their dyadic behavior in terms of patterns of interaction (Storch, 2002). Although no significant differences could be found intermodality, the tendency was that learners in the oral and written tasks generated more LREs than when only speaking, and they showed certain indication of a more collaborative pattern. Finally, the focus of those LREs was on average more on lexis than on form.

Task modality has also been recently investigated in terms of learners' previously known languages (PKL) use and their functions. For instance, Martínez Adrián and Arratibel Irazusta (2020) examined the interaction of 50 Basque/Spanish L1 child learners (ages 10-11) in an English CLIL context while performing two different tasks: (i) an only oral picture-rearrangement and narrating task and (ii) oral + writing decision-making task, similar to the one employed by García Mayo and Imaz Agirre (2019). Their findings showed that learners in general relied quite often on PKL to perform the task, but they did so to a greater extent in the case of oral + writing task. Nevertheless, the authors did not find an impact of modality on the functions of those PKL, as generally learners tended to use them for metacognitive functions and vocabulary searches. On the contrary, grammar-related talk in the PKL was minimal in both task modalities.

A second strand of research and pedagogical decisions is related to task implementation variables. One of the most researched factors has been task repetition (TR) (Bygate, 2009, 2018). Since learners' attentional and processing resources are limited and selective, it is considered that they cannot attend to both form and meaning simultaneously. According to Bygate, the first time that learners perform a task, they will be more inclined to concentrate on the task meaning and outcome than on form (that is, getting the message across and completing their task duty). Conversely, the second (or subsequent) time, learners will be able to resort to their experience and memory of their first performance and, in addition to being more fluent, they will devote more attentional resources to grammar and morphosyntax, hence producing a more complex language. TR in child TBLT has been studied in relation to its impact on attention to form and negotiation of meaning, L1 use, patterns of interaction and the complexity, accuracy and fluency triad (CAF, explained below).

Regarding TR and negotiation of meaning, one of the earliest studies was that by Pinter (2007a). In the context of the study reported above (Hungarian EFL, 10 years old, low L2 proficiency), she recorded a dyad who carried out a spot-the-differences task three times (using a different set of pictures, that is, using procedural repetition). She observed that children provided assistance to each other and paid attention to each other's utterances in successive times. Therefore, the author suggested that TR could work effectively with children of this age and proficiency.

More recently, there have been two other studies in the Spanish EFL context. García Mayo and Imaz Agirre (2016) analyzed the NoM strategies of 120 EFL children in a CLIL program, some of which were in the 3rd primary year (mean age 7.9) and other in their 4th year (mean age 8.89). At Time 1, they all completed a spot-the-differences oral task, but at Time 2 they were divided into three subgroups depending if the performed (i) same-task repetition (11 dyads in both 3rd year and 5th year), (ii) procedural task repetition (5 dyads in 3rd year, 11 dyads in 5th year), or (iii) a completely different task (the comparison group, consisting of 11 dyads in 3rd year and 12 in the 5th). Yet, the researchers did not find any difference regarding the children's NoM between Time 1 and Time 2 in any of the treatment groups.

The second study was carried out by Lázaro Ibarrola and Hidalgo (2017), also in the context of CLIL in Spain. They analyzed NoM as a result of procedural repetition in a picture placement task. The participants were 10 pairs of children (11 years old), and in contrast to the previous study, they completed the tasks three times. Their findings showed that at Time 3, the amount of confirmation checks and repetitions decreased significantly, which led the authors to suggest that TR could not be of help if the goal was to promote NoM among YLs.

Another variable studied in relation to TR has been L1 use. For instance, Azkarai and García Mayo (2017) analyzed the oral interaction of 42 Spanish YLs (ages 9-10). These learners had to carry out a spot-the-difference task at two times. While some learners performed same task repetition, others followed a procedural task repetition. Their unit of analysis for the children's interaction was c-unit (Foster et al., 2000) and the L1 use was classified according to the function it fulfilled: clarification request, confirmation check, lack of knowledge, phatics, repetitions, metacognitive talk, appeal for help and borrowing. Their findings showed that at Time 2 the use of L1 decreased in both experimental conditions, while the variety of functions stayed equal (for searching vocabulary and borrowings, most of the times). The authors link this preference of the L1 for such uses with YLs' willingness to avoid communication breakdowns and to make language meaningful.

With respect to the study of TR and patterns of interaction (Storch, 2002, 2009a), the study by García Mayo and Imaz Agirre (2016), explained previously, investigated this. They examined whether repeating an oral task procedurally or identically at two testing times impacted on child leaners' dyadic interaction, taking into account their age group (7-8 vs 8-9 years). In third-year primary (the lower age group), the majority of dyads exhibited at Time 1 a collaborative pattern (12 out of 16), and the rest were identified as passive/parallel. However, at Time 2 those four dyads turned into a collaborative pattern either performing procedural or same-task repetition. In fourth-year primary, conversely, at Time 1 most dyads displayed a passive/parallel pattern (17 out of 21). Yet, at Time 2 same-task repetition only helped one dyad to form a collaborative pattern, while procedural TR proved more successful, as more than half of the dyads in that condition turned to collaborative.

More recently, Azkarai et al. (2019) also explored the relationship between TR and dyadic patterns of interaction, in this case, in the Australian ESL context. Fourteen dyads (ages 6-8) with a low-intermediate level in English performed an oral spot-the-differences task twice and their interaction was classified in terms of equality (i.e. degree of control over the task) and mutuality (i.e. engagement with each other's contributions). The wide range of patterns found in this study contrasts with the previous one, as the authors identified up to five different patterns. The most frequent at both testing times was the cooperative pattern (7 out of 14 dyads at Time 1 and 6 out of 14 at Time 2), which is characterized by a very limited engagement from both partners, including short questions and answers, like the following example (Azkarai et al., 2019, p. 11):

(4)

J:	do you have two hills and blue sky?
D:	no.
J:	your shot.
D:	do you have a frog that's in front of the water?
J:	no.
D:	your shot.
J:	do you have? Do you have any water?
D:	yes.

In example (4), we can see that both learners respond to each other's questions about the elements of the picture with monosyllabic answers, very different from the spontaneous and imaginative interactions found in other studies with ESL children and more typical of collaborative patterns (Philp et al., 2008). The second most frequent pattern at Time 1 (4 out of 14) and Time 2 (3 out of 14) was collaborative. The rest of the patterns included dominant/passive, expert/passive and dominant/dominant. In general, the authors did not find TR to be a determining factor in establishing patterns of interaction, since only four dyads experienced a change and, in any case, those changes did not involve an increase in mutuality, and only two pairs increased their equality.

Finally, the impact of TR on the complexity, accuracy and fluency (CAF) of children's interaction has also been studied. For example, Sample and Michel (2015) analyzed six EFL learners' (mean age 9.5) performance in a spot-the-difference oral task, which took place three times. YLs' oral proficiency ranged from advanced beginner to lower intermediate. TR appeared to have a positive influence on task performance, but the CAF dimensions did not experience the same development. In general, the authors found a trend for an increased fluency at the expense of more accuracy errors throughout repetitions. These opposing directions led the researchers to believe there was some evidence to support Skehan's Trade-off hypothesis (2009), according to which, due to limited attentional capacities, a greater attention to a linguistic area may provoke a negative impact on another. Similar evidence was found by Bret Blasco (2014). In her longitudinal study, she studied the oral performance of 52 learners (ages 9-10 at the onset), 20 following traditional EFL and 32 attending CLIL. Oral data was gathered by means of an interview, but also through a picture-narration task, at four points in time. Although syntactic complexity and fluency gradually increased over each time, the opposite was true for accuracy.

Having a larger sample of child Spanish EFL learners (n = 120, age groups 8-9 and 9-10), García Mayo et al. (2018) examined the impact of procedural (16 dyads) and exact TR (21 dyads) on oral CAF. The task chosen in this case a spot-the-differences task, which the learners completed twice, with a three-month gap between each treatment time. In addition, a control group (23 dyads) carried out a completely different task at Time 2 (namely, a picture description task). The authors only found significant gains in the procedural TR group of younger learners' fluency and older learners' accuracy. With regards the rest of the measures, although no significant differences were found between Time 1 and Time 2, the younger learners decreased their rates in lexical and syntactic complexity. Thus, to a certain extent their findings echo the ones from the two aforementioned studies, as they showed that TR has a different impact regarding each CAF dimension. Indeed, familiarizing children with task procedures seem to positively influence to a greater extent their fluency and accuracy, but not their complexity.

The last implementation variable in relation to child task-based performance is the pair formation method. Although this variable has not been as extensively studied in the literature as TR, there are reasons to believe that it can moderate child learners' task performance, as has already been shown with adults (Mozaffari, 2017).

As mentioned above, especially when referring to the affective factors surrounding TBLT, interpersonal relations begin to gain importance towards late childhood, since YLs start paying more attention to the concept of friendship. As a result, they are more aware of who they feel comfortable working with and show a preference for working with their friends. However, the criterion of friendship tends to be less important when teachers are to form pairs, as they usually base their decision on the learners' academic level or even personality traits (Harmer, 2013). In fact, the difference in expertise between learners of the same dyad is also relevant from a sociocultural point of view. SCT theory initially considered that expert-novice interaction can lead to scaffolded assistance, the only type of assistance considered effective for learning according to Wood et al. (1976). Nonetheless, empirical sociocultural TBLT research has shown that similar proficiency pairs can engage in collective scaffolding, that is, learners can pool linguistic resources and perform beyond their individual capacity provided they collaborate and share a collective mindset (Donato, 1994; Sato & Viveros, 2016; Storch, 2002).

The study by García Mayo and Imaz Agirre (2019) examined the variable of task modality, but also took into consideration the pair formation factor. In this study (n = 62), twelve dyads were researcher-selected (those learners who had obtained similar scores in the proficiency test were paired together), eight dyads were teacher-selected (based on the teacher's knowledge of children's personality) and eleven dyads were self-selected (children worked with whom they considered their friends). The results showed that researcher-selected pairs produced more LREs in both task modalities (oral and oral + written), followed the teacher-selected pairs and, finally, learner-selected dyads. Therefore, their findings suggested that pre-existing friendship can generate more off-task behavior.

Basterrechea and Gallardo del Puerto (2020) also examined the variable of pair formation in relation to LREs and patterns of interaction. Fifty-four learners in a CLIL setting from the 5^{th} and 6^{th} primary school years (ages 10-12, L1 Spanish) were randomly assigned to two conditions: proficiency-matched dyads (i.e. researcher-selected on the basis of the children's proficiency scores in a standardized test) and self-selected dyads. Altogether, there were in 17 pairs in the former condition and 10 in the latter. They completed two consecutive collaborative tasks, one with an oral outcome and another with a written one, but in both cases learners had to negotiate for meaning in order to reach an agreement. Yet, the analysis of the LREs was reported together, as task modality was not an independent variable in their study. Their findings revealed that, regardless of the pair formation method, child learners produced more meaning-based than form-related LREs. What is more, in the case of meaning-related LREs, no statistical differences were found between the two groups with regards to the target-likeness of these episodes (which were largely correctly resolved). The similarity of the two groups was also present in the comparison of form-based LREs' target-likeness. Therefore, the researchers concluded that both groups performed alike, although they could find a tendency for researcher-selected dyads to generate more FonF and be slightly more accurate in their resolutions. However, as far as patterns of interaction is concerned, researcher-selected pairs displayed more collaborative engagement than their self-selected counterparts, who tended to be more dominant-passive or dominant-dominant.

3.2.3 Conclusions

After reviewing the main theoretical underpinnings of TBLT with regards to the characteristics of YLs, we have found support for idea that a task-based approach to language learning suits the child L2 learner population. These reasons include the prevalence of incidental learning strategies, the promotion of interaction and collaboration, the emphasis on learning-by-doing or experiential learning philosophy (Dewey, 1997), which encourages learners' engagement, and finally, the development of interpersonal relationships which are linked to affective factors that mediate cognition. Nonetheless, difficulties have been reported when implementing TBLT in primary schools, due to practicality matters (e.g. the impossibility of maintaining discipline in large classes) and also to teachers' misconceptions about this framework (Carless, 2002, 2003, 2004). These reservations are now being confronted with a growing body of empirical research that explores how tasks can be adapted to children's needs. The two main branches in this field have been (i) the comparison between adults' and children's negotiation of meaning in interactive tasks and (ii) the analysis of the impact of a wide range of task variables on YLs' performance. It is crucial that research continues informing TBLT pedagogy in primary school settings as well as broadening its scope. Research avenues include exploring the role of affect more in depth (Butler, 2017) and the use of new technologies, which are extremely popular with YLs (González-Lloret, 2017).

In the present dissertation, the use of a collaborative writing task, the dictogloss task, to help older children (aged 11-12) to focus on formal aspects of language is proposed. The next section will review the literature on collaborative writing (CW) tasks and the writing-to-learn approach. This type of tasks have not been so extensively used with YLs (Zhang & Plonsky, 2020), hence the need to shed more light on their effectiveness. In fact, as we can conclude from the review of empirical studies, oral tasks and input-based tasks have dominated the scene, with a few exceptions (Basterrechea & Gallardo-del-Puerto, 2020; García Mayo & Imaz Agirre, 2019; Martínez Adrián & Arratibel Irazusta, 2020). Therefore, in the next section, while summarizing the general CW research, we will highlight those pieces of research that have employed this sort of tasks with children.

CHAPTER 4 Collaborative writing tasks

In the review of the different paradigms that inform TBLT, it was indicated that from the cognitive-interactionist point of view, task modality refers to whether a task is oral or written, and this is a variable considered to influence learners' task performance. Nevertheless, interest from SLA researchers in writing does not have a long tradition, and this is reflected in the main TBLT constructs, which are related to the oral mode (e.g. negotiation for meaning or *languaging*) and which take writing as a subsidiary aid to speaking (Byrnes & Manchón, 2014). In fact, Byrnes and Manchón (2014) contend that much of the body of TBLT empirical research which has focused on writing was driven by the wish to ascertain whether those constructs fitted "the process and product of composing" (p. 4) rather than by a genuine attempt to understand writing itself or the process of L2 learning through writing.

As Manchón and Williams (2016) explain, the main reason for this neglect lies in the divergent objectives, and even epistemological and ontological principles, underlying two fields which have not been connected until very recently: L2 writing and SLA. On the one hand, research in the former field, especially the one taking place in ESL settings, has prioritized genre conventions, writing strategies and metacognition instruction over language instruction. This lack of interest in language development recently led to a call for more focus on language in writing classes (Polio, 2019). In fact, Polio suggests that the absence of a linguistic component in composition classes might be one of the reasons why different pieces of longitudinal research on L2 writing failed to show any improvement in the areas of complexity and accuracy (Roquet & Pérez-Vidal, 2015; Serrano, 2011; Storch, 2009b).

On the other hand, and of relevance for the present dissertation, scholars working in SLA aim to find generalizability in their results, and since writing is dependent on a myriad of factors (literacy, opportunities for practice, etc.), it was not considered useful. It should be noted that in SLA the idea of obtaining spontaneous learner production which depicts interlanguage as realistically as possible has dominated. Hence, writing, a mode which favors reflection and monitoring, has not been deemed as convenient as speaking. Nonetheless, thanks to the efforts of different scholars (Manchón, 2011; Ortega, 2012; Polio, 2012; Williams, 2012), in the last decade there has been a growth in the body of cross-disciplinar empirical research.

In a meta-analysis carried out from the studies published in the *Journal of Second Language Writing,* Riazi et al. (2018), after analyzing three periods of time (1992-1999; 2000-2009 and 2010-2016) reported that there was an increase in the proportion of research devoted to the theme of “language and literacy development” (from 5.4% in the first period to 16.5% in the third one). Conversely, the theme of “L2 instruction” slightly declined (from 16.3% to 12.1%). In other words, this contrast in trends highlights an emphasis in the “L2” component of L2 writing. Moreover, the theoretical paradigms which underpin L2 research have also been diversified. As the aforementioned meta-analytic study shows, in the first period of time it was the cognitive strand that dominated (18.9% of the total), that is, research was more concerned with the beliefs and mental processes influencing writing. Yet, research on the social aspects of writing, that is, the mediating social contexts and practices, were at that time very scarce (4.1%). In the third period of time, although the cognitive paradigm still prevailed (20.2%), the social one soared to this exact same proportion. What is more, studies merging both strands were also present in the corpus (accounting for 8.3%). Collaborative writing (CW), as one of the subfields of the broad field of L2 writing, was no exception to this increase. In fact, taking the field of applied linguistics as a whole, the number of studies related to this topic increased (although not significantly) over the period 2005-2016 (Lei & Liu, 2019).

Riazi et al.’s (2018) study, however, also called for attention to the fact that L2 writing research needs to broaden the scope of education contexts. For instance, in the last period of time that they studied (2010-2016), ESL contexts still prevailed over EFL contexts (53.7% over 29.4%), and higher education contexts over school contexts (52.2% over 8.8%). This is even more surprising taking into account that previous meta-analyses (Lee, 2016; Ortega, 2009) had called for research in school settings. Moreover, in a recent systematic review of face-to-face CW, Zhang and Plonsky (2020) also reported that research on children was still underrepresented. Hence, in the present dissertation we want to contribute to the still scarce body of early L2 writing-SLA connection by examining in particular YLs and CW. In what follows, we will review the main cognitive/psychological and sociocultural principles that underlie the use of CW tasks to develop L2 knowledge.

4.1 The contribution of L2 writing to L2 development

In Manchón and William's (2016) review of the interconnections between the fields of L2 writing and SLA, the contribution of L2 writing to L2 proficiency is considered one of the most promising research avenues. Different cognitively-oriented and sociocultural insights inform the study of the language learning potential of L2 writing (Manchón & Vasylets, 2019). Starting from the cognitive strand, the potential of writing to become an L2 learning site derives from three main factors inherent to this mode, summarized by Manchón and Byrnes (2014, p. 5): (i) the availability of time, (ii) the visibility and permanence of record and the possibility of obtaining feedback, and (iii) the problem-solving nature of the writing activity.

Thanks to the fewer time constraints in writing (with the exception of computer-mediated writing contexts, such as synchronous chats), it has been claimed that learners are able to better control their attentional resources. Consequently, they are more likely to reflect on their language use and resort to different knowledge stores, which eventually allow them to edit and monitor their output (Williams, 2012; Zalbidea, 2021). Moreover, revision and monitoring of output are also made possible by the permanence of record. Regarding the different kinds of knowledge stores available while writing, cognitively-oriented scholars have proposed two different views. On the one hand, it is suggested that, given the extra time available, learners may access explicit (conscious) knowledge, which after subsequent attempts, ultimately promotes implicit learning. In other words, L2 writing would restructure L2 knowledge from explicit to implicit (N. Ellis, 2011). On the other hand, other scholars contend that writing leads to an automatization of explicit knowledge (as a consequence of drawing constantly on it), that is, without generating any qualitative change in the L2 knowledge characteristics (DeKeyser, 2007). These two views of the type of knowledge that writing promotes are crucial for TBLT as well, as we have already seen that the cognitive account generally argues that incidental learning should be fostered within this framework (R. Ellis et al., 2020).

Finally, another psychological advantage of writing over speaking is related to the problem-solving nature of the former, which is claimed to trigger learning processes that facilitate L2 acquisition, such as noticing and metalinguistic reflection by means of FonF. As far as noticing is concerned, it is predicted that when learners struggle to meet the communicative requirements imposed by the task, they may become aware of what they are not able to express and would like to be able to express. This process was labelled by Swain (1998) as "noticing holes" in the learners' interlanguage. Noticing one's own interlanguage deficiencies differs from other kinds of noticing described by Swain (1998), such as "noticing the gap" (i.e. when learners become aware of the differences between their interlocutors' and their own production) or noticing a form in the input (i.e. when learners simply attends to formals aspects of the language while).

As different studies have demonstrated in the case of written output (Hanaoka & Izumi, 2012; Uggen, 2012), when this process occurs, learners begin to pay more attention to the input they receive in order to fill in those holes, which in turn shapes intake (i.e. processed and registered input in learners' L2 system). Furthermore, although this process might also occur in speaking, writing allows learners not only to seek responses in the input, but also to fall back on different resources, such as consulting experts or peers, reference materials and processing feedback[2]. More importantly for the present dissertation, it has been demonstrated by a large body of empirical research that learners may not only benefit from feedback in L2 writing provided by teachers or experts, but this is also true for peers (for a review, see Yu & Lee, 2016).

Regarding focus on form, despite the fact that both speaking and writing processes may facilitate these cognitive moves, the latter is considered to offer a greater opportunity in terms of time to reflect on language use, and it provides the chance to plan and monitor production. In fact, as Schoonen et al. (2009) explain, *a priori* the written register is less tolerant towards errors and demands a greater level of accuracy than speaking. Moreover, being able to see the text also should in principle make learners aim higher, and produce "pushed output" (Swain, 1995, 2005), that is, output

[2] Although the study of written corrective feedback (WCF) has been a major theme in the L2 writing literature (Manchón & Vasylets, 2019; Manchón & Williams, 2016), it is not addressed in this review as it goes beyond the scope of the present dissertation.

which, apart from accurate, is also precise, coherent, cohesive and appropriate. However, studies comparing task-modality production have shown mixed results in this respect (Manchón & Williams, 2016).

4.2 Collaborative L2 writing: A sociocultural justification and review of the main empirical research

The capacity of L2 writing as a site for focus on form and linguistic reflection has, indeed, been claimed to be increased when it involves collaboration. Collaborative writing (CW) has been long studied by scholars working in pure composition studies (Ede & Lunsford, 1992), who link this activity to real life situations in workplaces in which writing usually takes place in teams (Storch, 2019a). Nevertheless, it has also received attention from authors in SLA, who, starting with the work by Swain and her associates (Swain & Lapkin, 1995, 1998), have concentrated on its language learning potential from a *writing-to-learn* approach (Manchón, 2011). As Zhang and Plonsky (2020) concluded in their recent systematic review of face-to-face L2 CW research, this rather new field (it only started to take off at the beginning of the 2000s) draws on "L2 writing (e.g. CAF, metrics of text quality), shares the research orientation (i.e. classroom-based) and analytic methods (e.g. LREs) with task-based L2 research, and probes into constructs (e.g. uptake, L2 learning gains) that are typically examined in SLA research" (p. 15). Despite this eclecticism, the same review clearly shows that most research on collaborative L2 writing has primarily relied on SCT to explain its learning potential (especially, to interpret learner interaction), and therefore, we will start by explaining the main sociocultural constructs which justify adding a collaborative component to L2 writing.

Firstly, we will define what we understand by collaborative L2 writing. One of the scholars that has most extensively studied collaborative L2 writing theoretically and empirically from an SCT perspective has been Storch (1999 *et passim*). In her TESOL encyclopedia entry for the term 'collaborative writing', she defines it as "the coauthoring of a single text by two or more writers, where the coauthors are involved in all stages of the composing process and have a shared ownership of the text produced" (Storch, 2018). This way, the author distinguishes collaborative writing from other modes of writing, such as cooperative writing, where learners co-contribute to the text elaboration (for instance, with division of labor, such as sharing the tasks), but do not co-construct it (a term which will delve into further).

According to sociocultural tenets, knowledge and human cognition are socially constructed by members of the community (Dewey, 1997; Vygotsky, 1978). Language is considered a mediating tool that regulates mental and physical activities. Just as in other fields of knowledge language is used to grasp higher-order concepts (Lantolf, 2000), and in SLA language constitutes both the means and the object of study. Thus, when composing a text in collaboration, the interaction that takes place between the learners gives way to assist each other, and when such assistance is within the learners' ZPD it is referred to as *collective scaffolding* (Donato, 1994).

As Storch (2016) explains, the application of SCT principles to explain SLA processes led to the reformulation of the Output Hypothesis by Swain (2000), who shifted her interest from the analysis of the output itself (be it written or oral) to the actual process, that is, the communicative and cognitive activity generated between the learners. Swain coined the term 'collaborative dialogue', and later, 'languaging' (2006) to refer to the "process of making meaning and shaping knowledge and experience through language" (p. 98). Arguably these problem-solving deliberations may help co-construct knowledge (gaining new knowledge or consolidating previously acquired one). In empirical research, and this includes TBLT studies focusing on learners' linguistic deliberation during CW task performance, languaging has been operationalized in the form LREs (Swain & Lapkin, 1995, 1998).

Collective scaffolding and languaging have been argued to be forms of feedback just as valuable as the well-researched written corrective feedback (usually provided by the teacher or the researcher) (Bitchener, 2008, 2016; Bitchener & Storch, 2016) or feedback provided by peers during multiple drafts (Hansen & Liu, 2018; Hansen, 2005). As Storch (2019b) explains, peer feedback during collaborative writing tasks shares some characteristics with the latter, such as the fact that feedback is provided on all aspects of writing (that is, not just concerning superficial errors) and that the feedback is accessible in sociocultural terms (it is calibrated to suit learner's needs). We provide two examples from our database that display each of these feedback features:

(5) [MAR001 & MAR014 - Day 2]

*CHI B: *¿Qué has escrito?* celebration? (What did you write? Celebration?)
*CHI A: *No, pero que no pongas el título da igual* (But don't write the title, it's ok).
*CHI B: *A ver sí, el título es lo más… tenemos que hacerlo* (Of course we have to! It's the most… We must write it).
*CHI A: A celebration.
*CHI B: *Pues venga (…) ¿Qué haces? Ponlo en el medio* (Ok… What are you doing? Write it in the middle).
*CHI A: *¿Qué más da?* (It doesn't matter!)
*CHI B: *No, sí más da* (It does matter).
*CHI A: *Ya está, ¿contento?* (Fine, happy?)

(6) [MAR 009 & MAR012 - Day 1]

*CHI A: She is hungry and she eat her…or his… her or his, his his!
*CHI B: *No, es* 'her' (No, it's 'her').
*CHI A: 'His'.
*CHI B: *Es* 'her', 'his' *es para chicos*… (It's 'her', 'his' is for boys)
*CHI A: *Tengo un lío ahí…* (I have a mess there)
*CHI B: *A ver,* 'her' *es para chicas y* 'his' *para chicos. Yo antes ponía* 'his' *para todo* (Let's see, 'her' is for girls and 'his' is for boys. I used to say 'his' for everything before).

In (5), the pair is discussing a topic beyond language form. We can see that CHI B insists on adding a title to the story, while CHI A does not consider it necessary. Once CHI A compromises, he notices that the layout of the title is wrong and requires his partner to change it to the middle, despite the fact that CHI B does not think this is important. Therefore, each of the learners gave and received feedback at a discourse level by pooling resources together and sharing their individual understanding of the task requirements. In (6), CHI B provides her partner with feedback on the third person singular possessive determiners (POSS, one of the target forms in the present dissertation, as we well explain below). Although CHI B takes the role of expert and CHI A of novice ("I have a mess there"), CHIB does not use of technical metalanguage (the term 'determiner' is not used) and she even explains that she used to have the same problem (i.e. showing some empathy towards CHI A). Hence, this feedback features suggest that the explanation is developmentally appropriate for CHI A. Storch (2019b) argues that the feedback provider also benefits from this dialogic interaction, since by "articulating these explanations [s/he] can deepen their own understanding of language conventions, and of form-meaning connections" (p. 153). Moreover, the use of the L1 in both examples indicates a disposition from the learners to be closer to one another despite having a different knowledge of the L2.

Although, as we have mentioned, these two features of feedback can also be found in peer-response activities, Storch (2019b) indicates two features which are unique of CW. The first is the fact that feedback is contingently responsive to the learners' needs, that is, it is provided exactly in the moment when an error is noticed or there is a request for assistance. In contrast, according to Storch, peer-response activities lack this feature, as feedback is generally provided on the basis of previously elaborated checklists, which may disregard the actual learners' needs. The second feature is that feedback is motivated. This characteristic is related to the fact that the responsibility over the final text is shared by the co-authors, and thus, they are more likely to provide feedback in order to achieve the best possible outcome. Conversely, in peer-response activities, since the student who edits a text is not its owner, they are less willing to invest themselves into this endeavor, and sometimes they might even be reluctant to do so for fear that they will cause offence.

To conclude this section, we will review some empirical studies that have demonstrated that CW provides more language learning and FonF opportunities than solitary writing. In fact, this has been the most frequent independent variable (IV) in quasi-experimental CW research (present in 27% of all published studies from 1992 to 2017) (Zhang & Plonsky, 2020). Once again, we must start by mentioning the pioneering work by Storch in the Australian ESL context. In one of her earliest studies (Storch, 1999), she compared how dyads and individuals performed a series of tasks; one of them had a clear grammatical focus (a cloze text), while the other two were closer to writing tasks (a composition and a text reconstruction). Her findings showed an overall positive effect of collaboration on grammar accuracy, although the effect depended on the linguistic feature. In a later study (Storch, 2005), the author also compared intermediate ESL learners' writing in pairs (n = 9) and individually (n = 5). The positive effect of collaboration was not only perceived in the quality of the writing (as dyads produced more accurate and complex, albeit shorter, writing), but also in the fact that pairs had the opportunity to scaffold each other and pool their knowledge.

The impact of collaboration on the learning potential of L2 writing has also been investigated in FL settings with adult learners, and in some cases, the findings did not converge with those from ESL. For example, Shehadeh (2011) conducted a longitudinal study (16 weeks) with university students in the United Arab Emirates. While 18 students carried out the writing tasks in pairs, 20 completed them individually. Using a qualitative rubric, their written product was assessed in a pre- post-test design. The findings showed that some linguistic areas (content, organization and vocabulary) were positively influenced by collaboration, whereas grammatical accuracy was the same for both groups.

Watanabe (2019) compared how EFL university learners in Japan (n = 20, ages 18-20) performed two writings: one individually and another one in pairs, on two different days. Innovating on previous research, she also recorded the collaborative dialogue in pairs, she tried capturing the languaging of individuals by recording what she termed "speech for self". Their written performance was assessed on the basis of an analytic rubric. Her findings failed to show any advantage of collaboratively produced texts over individual ones. Furthermore, the results showed that the benefits of collaboration were only present when learners had a collaborative pattern during the task (those dyads scored the same or a higher score as in individual writing), but the case was the opposite for uncollaborative pairs.

How the number of learners involved in collaboration is more or less beneficial than individual writing has also been explored in the literature. Fernández Dobao (2012) was the first study to compare group, pair and individual work in CW tasks in the L2 classroom. The study was conducted with six intermediate classes of Spanish as a foreign language (SFL) in the USA. Twenty-one learners worked individually, thirty in pairs and sixty in groups of four on a jigsaw task. They had to rearrange the pictures provided and produce a written text. Fernández Dobao examined whether the number of participants had an effect on the fluency, complexity and accuracy of the written products and on the frequency and nature of the oral interaction produced in pairs and groups. Regarding the former, she reported that the texts written by groups were not only more accurate than those written individually, but they were also more accurate than those written by pairs. As far as oral interaction was concerned, there were no differences in the number of instances devoted to form by pairs and small groups, but the latter were able to correctly resolve more LREs than the former. The authors justifies the advantage of small groups over pairs with the greater number of resources (i.e. knowledge of the members) participants could pool during the task.

More recently, also in an SFL context, Gallego (2019) discussed the effect of two collaborative dictogloss tasks (to be further explained) on the subjunctive recognition and written production by 104 learners at a university in the US. Participants were assigned to two treatment groups: dictogloss (n = 33) and dictogloss + explicit instruction (n = 35). The control group (n = 36) did not engage in collaboration or writing, but instead carried out a series of meaning-focused activities that aimed to trigger the subjunctive use. During the pretest, posttest and delayed posttest, all participants had to carry out a writing task and a subjunctive recognition task. The findings showed that the two treatment groups were able to increase their target-form production at similar rates, and, indeed, the gains were higher than for the control group. Moreover, regarding text quality, in the post-test and delayed post-test the treatment groups produced significantly longer texts than the control group. However, no intergroup difference was found with regards to text complexity. In sum, Gallego argues that, irrespective of the explicit instruction, by merely working collaboratively, learners benefit from metacognition and metatalk.

As with L2 writing research in general, little research on collaborative vs individual L2 writing has been carried out in school settings. One of the first studies was conducted by Kuiken and Vedder (2002b), from an interactionist rather than a SCT approach. In the Dutch EFL setting, they compared how performing a dictogloss task twice in pairs (n = 20) and individually (n = 14) could help adolescent learners (ages 16-18) in the acquisition of the passive structure in English (measured by means of a recognition task). Contrary to what was shown by Storch (1999; 2005), they could not find evidence to support that the opportunity to interact with peers was more beneficial than individual text reconstruction with regards to passive form identification, but they did find that interaction "stimulated noticing" (p. 354).

Basterrechea and García Mayo (2013) investigated the effects of collaborative work on the production of the present tense marker *-s* by 41 CLIL and 40 EFL learners (age range 15-16, L1 Spanish) during dictogloss. Collaborative text reconstruction led to more accurate use of the target form than individual text reconstruction in the two educational contexts, and CLIL dyads who collaborated outperformed those who worked individually in the same setting.

In a recent study with adolescent EFL learners (age range 16-17, intermediate proficiency level), Villarreal and Gil-Sarratea (2019) explored the learning affordances of CW. In their study, a control group (n = 16) produced an argumentative text individually and an experimental group (n = 16) did so in pairs while their interaction was being recorded. Their findings were congruent with previous research, as they revealed that pairs produced shorter but more accurate and slightly more lexically and grammatically complex texts. Moreover, texts were analyzed qualitatively using holistic measures and the findings showed that pairs also obtained higher scores in content, structure and organization.

Finally, the study by Calzada and García Mayo (2020a) investigated whether child EFL learners could benefit from a collaborative (in pairs or small groups of three) or individual dictogloss to speed up their acquisition of the third person singular -*s* marker (3S) and the articles (a, the, ø). In order to target these features, both elements were seeded through the text, which contained nine instances of each feature. Learners were randomly assigned to one of the treatment groups (pairs n = 16, small groups n = 16, and individuals n = 18) and they completed a GJT before and after the one-shot dictogloss task. Regarding 3S, pairs and small groups improved their scores from the pre-test to the post-test, whereas individuals' score decreased. In fact, the pairs' significant advantage over small groups, already present in the pretest, was maintained and increased in the posttest, where these learners also scored significantly higher than individuals. In the case of articles, pairs were the only ones to improve their score in the posttest, and the statistical test revealed significant differences between this condition and the other two treatment groups. Hence, the authors suggested that reconstructing a text in pairs might be more beneficial than solitary or small group dictogloss.

To summarize, the studies that have compared actual pair vs individual L2 writing have shown that collaboration generally increases the accuracy of the texts produced as a result of learner interaction and collective scaffolding during the task. In fact, in a recent meta-analysis reviewing studies that compared collaborative vs individual L2 writing (Elabdali, 2021), accuracy appeared as the only dimension where it could be confidently shown that collaboration brought about benefits as compared to individual writing. Apart from the quality of the written product, the opportunities for attending to form, quantified as LREs across different studies, seem to generate stronger changes in learners' knowledge about target forms as compared to individual writing. However, the scarcity of research looking into YLs' performance and the failure to account for the individual languaging that arises during individual writing clearly show that more research is needed to determine the benefits of collaborative L2 writing in school contexts. Finally, in the following section, the influence of CW on L2 writing and proficiency development is shown to be contingent on a series of factors, which, in general, can be grouped as (i) task-related and (ii) individual factors.

4.2.1 Variables impacting on collaborative writing tasks

4.2.1.1 Task-related factors

4.2.1.1.1 Task type: The dictogloss task

The usual classification of CW tasks distinguishes tasks which are meaning-focused and those which are language-focused (Storch, 2013, 2016). The former ask learners to write a text based on pictorial or written information. Their main goal is to create a meaningful text and, therefore, no grammar is preemptively targeted; instead, FonF arises spontaneously. Conversely, language-focused tasks are specifically designed with the aim to draw learners' attention to specific linguistic features. Among meaning-focused ones, researchers have used a wide range of tasks, from those closer to classical compositions (such as descriptive and argumentative compositions, or data commentary reports) to those which resemble more dialogic and information-gap task, such as the jigsaw task (Ashcraft, 2018). In a jigsaw, each learner receives some pieces of information which they need to share so that they can complete the whole passage.

Regarding language-focused tasks, there seems to be a continuum depending on the degree of explicitness by which learners' attention is drawn to form. On the more explicit and more controlled side, we can identify editing and text reconstruction tasks[3]. Although attending to the definition of task (R. Ellis, 2009) they cannot be considered as such (the only outcome is linguistic and the primary attention is not on meaning, to mention just two of the violated precepts), Storch (2013) explains that this term is widely used in the field to refer to these activities, although since they do not involve writing *per se* she does not include them in her revision of research on CW (2016, 2019a). In the former, learners are provided with a text created by them or the teacher where there are typical errors for their age and proficiency group which they need to correct. In the latter, the text provided contains gaps that learners need to fill in (usually inflectional morphemes or function words).

Conversely, on the more implicit and less controlled side of the continuum, we find the dictogloss task (García Mayo, 2018a; Wajnryb, 1990). According to Storch (2016), dictogloss has been the most widely used CW task in research. The canonical procedure involves listening twice to a short text at a normal speed. The text, designed ad hoc by the teacher or the researcher, is seeded with instances of a problematic feature for the learners, although they are not explicitly warned about this fact. During the second time they listen to the text, learners take notes of the key ideas individually. Afterwards, learners start working in pairs to reproduce that very same text. The main instruction is to produce a grammatical text which keeps the gist of the original (i.e. not a word by word resemblance is expected in the learners' writings, as in traditional dictation activities).

Meaning-based CW tasks, due to their open-ended nature and their inherent emphasis on content, tend to generate fewer LREs than grammar-focused tasks. Different studies comparing these two types of tasks support this claim. For instance, Storch (2001) compared 10 ESL adult pairs (with an intermediate proficiency in English) performing three different CW tasks: a meaning-focused task (text composition based on some data interpretation), and two grammar-focused tasks, including an editing task (based on previous students' reports) and a text reconstruction. The results showed that the grammar-focused tasks generated at least twice as many LREs as the meaning-

[3] In the literature "text reconstruction" has also been used to refer to the dictogloss task (Eckerth, 2008; Khezrlou, 2019), however, to keep consistency in this dissertation we will exclusively refer to the latter.

focused task, and what is more, the composition task induced learners to more deliberations about lexis than grammar, while the opposite was true in the other two tasks. Alegría de la Colina and García Mayo (2007) corroborated these results in their study comparing the effect of three tasks (jigsaw, text reconstruction and dictogloss). Twenty-four first-year undergraduate Spanish students of L2 English were assigned to groups of four and each of the groups performed one type of task. Although all three tasks proved to generate form-based LREs, the text reconstruction and the dictogloss, in this order, yielded more instances

Yet, there are also examples in the literature showing counterevidence. In a study by Swain and Lapkin (2001), the authors also explored CW task-type differences with adolescent learners (n = 65) in the French immersion context in Canada. In particular, they compared their languaging in a jigsaw and a dictogloss. Their findings did not show any significant differences in the number and theme of the LREs.

Moreover, research has shown that there might as well be differences within grammar-focused CW tasks. For instance, García Mayo (2002a) compared adult EFL learners' (n = 14) performance on five tasks: a cloze text, a multiple-choice gapped text, a dictogloss, a text reconstruction and a text editing task. The author found that the more explicit tasks (cloze text, multiple-choice, reconstruction and editing) generated more LREs related to language form and a higher proportion of LREs per turn, while the potential of the more implicit one (dictogloss) to draw learners' attention to form was more limited. What is more, despite the fact that each of the explicit tasks enabled learners to discuss their target forms to a certain extent, the dictogloss was the only task which did not generate any LREs on the target items.

As the dictogloss (DG) is the central task in this dissertation, we will now review more research analyzing its potential to draw learners' attention to form. One of the earliest examples is the study by Kowal and Swain (1994). In the French immersion context of Canada, nineteen adolescent learners (ages 13-14) completed a DG task four times. The analysis of their interaction showed that learners produced more grammar-related episodes than mechanical or lexical, and that they were able to provide each other with accurate feedback. Other studies, apart from quantifying and classifying LREs generated during DG, aimed at showing some evidence of learning. For example, LaPierre (1994), in the same context, demonstrated a correlation between the LREs the learners' resolved correctly and their performance at the tailor-made posttest designed by the researcher based on their discussions.

At this point we should highlight that one of the main problems involved in CW task studies is the difficulty in showing evidence of learning through tests which are previously designed by the researcher without taking into consideration the LREs (Storch, 2016). This problem is even more acute in the case of the DG task, as due to its implicit nature, it leads to open interpretations from learners about which the requirements of the task are. In fact, in line with García Mayo (2002a), more studies have shown that learners eventually discuss many other aspects which are not targeted by the task. Eckerth (2008), for instance, compared the linguistic gain of the target feature and non-target features made by adult L2 learners of German (n = 31) as a result of two collaborative DG tasks. In order to measure the gains on the target form (the passive voice) the author employed a pretest, posttest and delayed posttest designed *a priori* (without taking the actual learners' discussion into account). In order to measure the gains on non-target forms, Eckerth administered an *a posteriori* delayed posttest, where he tallied the correct answers that had been discussed during the task as "interaction-internal", and correct answers that had not been discussed as "interaction-external". The findings showed significant learning gains in the short and medium term on target features and a considerable contribution of the task to gains in non-target features. The author, however, posed a question as to which of the two gains would be more permanent in the long term, since "non-targeted L2 issues reflect the learners' alertness towards their individual learning problems, [and therefore] they might be of particular acquisitional potential" (p. 133).

For the purpose of this dissertation, we searched for published studies that used collaborative L2 DG (from 1994 until July 2020) in two databases, namely, Education Resources Information Center (ERIC) and Google Scholar. This combination of words were used: *dictogloss* OR *text reconstruction, collaboration* OR *collaborative work* OR *collaborative writing,* L2 writing OR *foreign language writing*. We retrieved 73 studies, but after careful examination, we selected and summarized thirty-five (k = 35). The main requirements that we set in order to include a study in our corpus was that (i) it should use collaborative L2 writing (including studies which used it as part of other tasks), (ii) the article was written in English, and (iii) it was published in a peer-reviewed source (PhD theses and unpublished MA theses were therefore excluded). The Excel file containing the summary of the studies was uploaded onto *NVivo 10* (QSR International, 2012) for coding their characteristics and to obtain a general picture of what had been (and what had not been) studied so far. The following table summarizes the most relevant context characteristics[4]:

Table 1 Collaborative L2 DG systematic review: participant and context characteristics

		k	%
AGE	Adolescents	11	31.4%
	Adults	21	60%
	Children (up to 12 years)	3	8.6%
L1	English	7	20.6%
	Spanish	7	20.6%
	Mixed	7	20.6%
	Other[5]	11	32.4%
	Not specified	2	5.9%
PROFICIENCY	Advanced	3	7.7%
	Intermediate[6]	19	48.7%
	Elementary[7]	8	20.5%
	High vs Low[8]	2	5.1%
	Not specified	7	17.9%
SETTING	EFL[9]	20	55.6%
	ESL	3	8.3%
	French immersion	4	11.1%
	Korean as a Second Language	3	8.3%
	SFL	3	8.3%
	Other[10]	3	8.3%

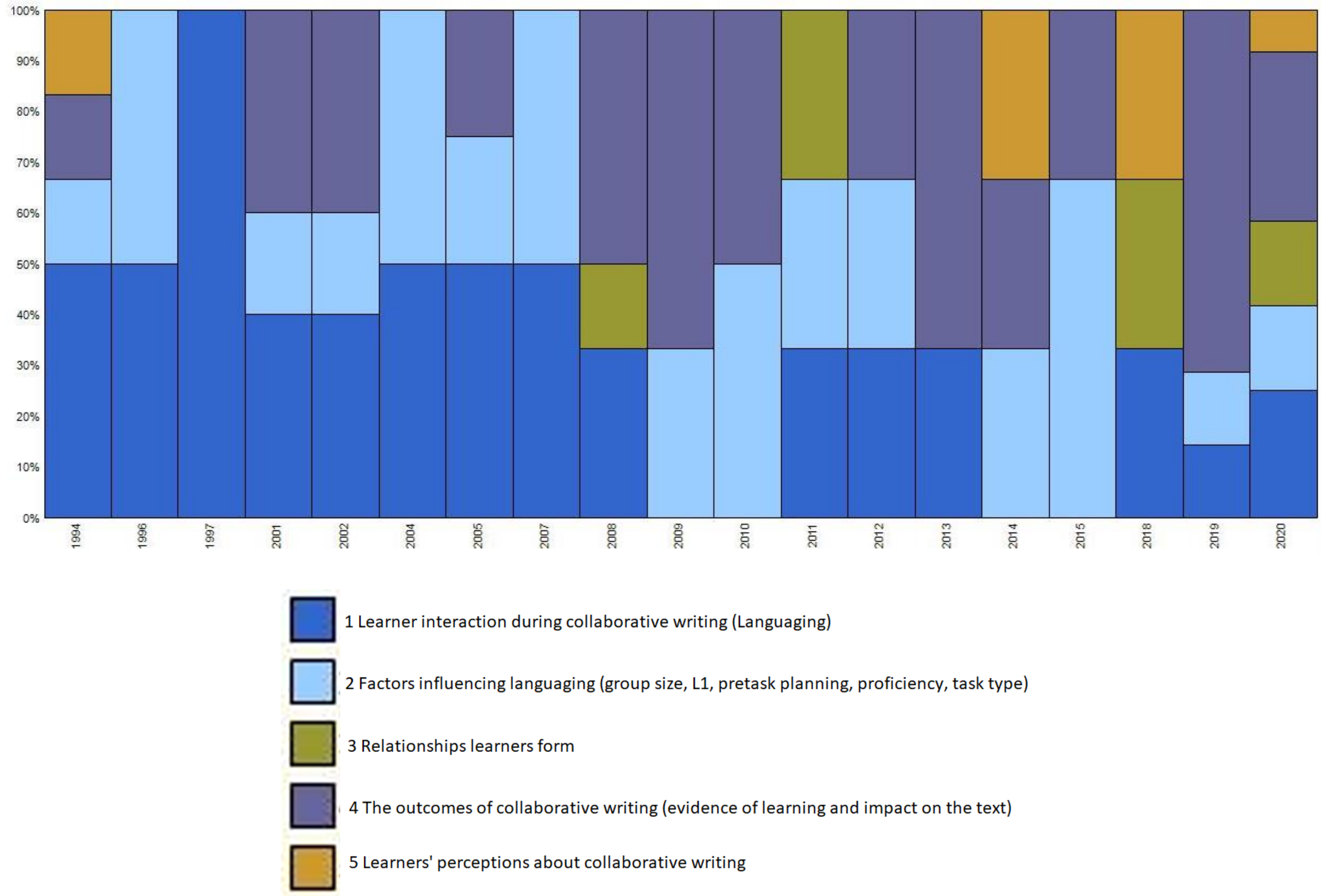
This systematic review shows some interesting figures. Regarding age characteristics, more than half of the studies have had adult participants, and only an

[4] Although the number of studies collected for analysis was k = 35, the total number in each of the categories can be higher, as some studies were coded into more than one (e.g. studies that had both advanced and intermediate participants).
[5] L1 Other included: Chinese, Dutch, German, Iranian and Swedish. Bilingual Spanish and Basque are counted within L1 Spanish.
[6] Including upper- and lower-intermediate.
[7] Including upper-elementary.
[8] Studies which split participants into two groups taking into account their scores in an in-house test, but without reporting their general L2 proficiency explicitly.
[9] CLIL in English included.
[10] Other included: Chinese as a Foreign Language and L2/L3 German.

8% has used the task with YLs. This trend mirrors the one shown in the systematic review on CW (Zhang & Plonsky, 2020), where 84% of the studies (1992-2017) examined adult learners. As for the L1, English and Spanish account for 41% of all studies, although the variety contained in "Other" should also be noted. This diversity means that DG has been used by researchers all over the world. In terms of proficiency, intermediate level is the norm. Yet, it is noteworthy that almost 20% of the studies did not report any information in this respect, just as Zhang and Plonsky (2020) found relevant to point out that 14.9% of the studies lacked this information in their review. Finally, the most frequent setting has clearly been EFL (55.6%), as the percentages of the rest of the settings lag far behind. This is also typical of CW studies, but not of general quantitative L2 research, where ESL settings are more common (Plonsky, 2013).

Regarding the research topics covered in collaborative DG studies, we organized the themes according to Storch's (2019) taxonomy (two topics, those related to computer-mediated learning and Activity Theory were excluded as we could not find any examples in the DG literature). Figure 3 displays this information:

Figure 3 Research topics in Collaborative DG studies across the years



As can be seen in the timeline, the interest in the topics mirrors the trend described by Storch (2019): the first period (in our case running until the beginning of the first decade of the new millennium) concentrated on the study of languaging (usually, in terms of LREs), as research on this topic constitutes 40% or more of all publications from those years. The second period starting from then is characterized, instead, by a wider range of research topics, emphasizing the impact of DG on learning and writing, the description of learner behavior during the task or even their perceptions about DG. To a certain extent, these two periods also reproduce the general trend in TBLT (Plonsky & Kim, 2016), with the first being more related to "task-as-process" and the second more linked to "task-as-treatment".

To conclude this brief review on collaborative DG, it was possible to quantify certain experimental design characteristics regarding task implementation which could inform the design of the present study. In fact, sometimes these decisions (e.g.

repeating the DG task or including a video-modelling session) were not considered as independent variables, but they were certainly confounding variables which could impact, for example, learners' focus on form compared to the findings from other DG studies which did not include them. This concern is shared by Zhang and Plonsky (2020), who suggest that "not implementing or reporting pretask training in CW makes it difficult to understand whether this critical factor might have influenced the findings, especially for quasi-experimental research that employs a one-shot design" (p. 13). The results are summarized in Table 2:

Table 2 Task implementation variables in collaborative DG studies

	k	% (over the total amount of studies, *k* = 35)
DG = 1	16	45.7%
DG + 1	19	54.3%
Pretask FFI	6	17.1%
Pretask modelling	9	25.7%
Posttask text modelling	2	5.7%
Corrective feedback	2	5.7%
Individual vs collaborative DG	6	17.1%

Note: DG = 1: DG performed only once / DG + 1: DG performed more than once

First, it is interesting to note that the number of studies where DG was implemented more than once was greater than the studies where it was implemented only once. Yet, the concept of task repetition is not investigated by any of the studies as an independent variable. Secondly, as DG is usually an unfamiliar task for learners regardless of their setting, almost a quarter of the studies included a video modelling session with the learners prior to the experimental stage. In this vein, given that the DG task is usually grammar-oriented, a few studies included a pretask FFI session where researchers, for instance, explicitly pretaught the target form with the aim of directing learners' attention towards it. With regards to post-task design, only a few studies included some sort of expert feedback (in the form of WCF or text modelling) to analyze learners' noticing during those stages. The comparison between individual and collaborative DG attracted the attention of several researchers. Finally, it is important to be aware of the difficulties involved in the comparison of the studies due to the different criteria regarding LRE classification or participants' proficiency measure.

As in the case of other CW tasks, researchers right from the start (Kowal & Swain, 1997) pointed out different factors which could impact the degree of attention to form during DG, some of them related to task conditions and others to learner factors. We will now explore how task repetition, a task implementation variable, can impact learners' performance in CW tasks.

4.2.1.1.2 Task repetition (TR)

The impact of task repetition (TR) has been mainly studied in oral tasks, especially with adult learners (Ahmadian et al., 2017; Bygate, 2009), and only more recently with YLs (García Mayo et al., 2018; Lázaro Ibarrola & Hidalgo, 2017). In the writing domain, the benefits of TR demonstrated in speaking (fostering FonF processes, drawing attention to more redundant features of production and, eventually engaging in deeper processing) have been argued to be equally relevant or even increased (by adding, for example, corrective feedback) (Manchón, 2014a). In fact, a few individual L2 writing empirical studies have sought to validate this theoretical prediction: a longitudinal study of 30 weeks, analyzed from a Dynamic Systems perspective (Nitta & Baba, 2014); a study in a computer assisted language learning environment comparing exact and procedural TR over the course of five weeks (Amiryousefi, 2016); and, finally, a study comparing TR alone and in combination with different forms of metalinguistic explanation (ME) across four DG tasks (Khezrlou, 2019). These three studies, all of them including adult learners, report linguistic benefits of TR in writing.

Yet, the combination of CW and TR is still an uncharted territory. In Wigglesworth and Storch's review article (2012), the authors already suggested that "repeated collaborative writing activities afford learners repeated practice in deliberating about their ideas, in giving and receiving feedback, and in rewriting" (p. 372), which eventually lead to language learning. In our previous review of the collaborative DG task literature, there was a high number of studies which contained a TR component, although the authors did not classify it as such or did not analyze it as a task condition variable (54.3% of all collaborative DG studies). In fact, on most occasions, TR was used to familiarize learners with DG. Nevertheless, the very recent studies investigating TR in CW tasks serve as an example of the scholars' interest in this research topic.

In adult FL settings, Kim, Choi et al. (2020) conducted a study with low-proficiency learners of Korean as Foreign Language (n = 54, ages 18-30, different L1 backgrounds), analyzing the combination of TR with synchronous indirect written corrective feedback over the course of 10 weeks. Their aim was to show the impact WCF and TR on the quality of the collaboratively written texts (in terms of CAF), on the knowledge of the Korean target form (Korean honorifics, assessed by means of a grammar correction exercise before and after the collaborative writing tasks) and their perception of collaborative writing after each of the two task enactments. For the purpose of this section, only those findings related to TR will be summarized, as WCF is out of the scope of the present dissertation. The results revealed that TR positively influence fluency, but it did not influence complexity or accuracy. The lack of effect on complexity (which is at odds with the general TR prediction) is hypothesized to be linked to learners' low proficiency in the language, as they still did not have a good command of complex structures. Moreover, the lack of improvement in accuracy is supposed to be due to "the nature of collaborative writing" (p. 19), which according to the authors diverts learners' attention towards meaning. Regarding task perceptions, TR positively impacted their views on task difficulty, task satisfaction, motivation for performing future tasks and perceived task performance. In contrast, the perceived learning opportunities slightly decreased from Day 1 to Day 2. Finally, after the two task enactments, learners were able to score significantly higher in the target form posttest.

In the same learning context (i.e. low-proficiency Korean as a Foreign Language learners), Kim, Kang et al. (2020) looked again into the interplay between TR and CW, but this time leaving out of the equation the influence of WCF. In this study, thirty-eight students engaged twice in information gap task which required writing a blog post in pairs: some of them experienced exact TR (n = 23), whereas others performed the same task with different content (i.e. procedural repetition) (n = 15). The most interesting aspect of this piece of research was its threefold line of enquiry, as the authors examined (i) task process (in terms of LREs), (ii) task product (the complexity and accuracy of the written production), and (iii) the learning outcomes of the target features. It should be highlighted that all students received a short grammar review of the target features, as well as a pretask modelling video. Regarding their attention to form, the findings showed a number of differences between the two conditions. In fact, procedural TR led to more target LREs on both days than exact TR. The former generated longer non-target LREs on Day 2 than the latter. In contrast to this, the exact TR group had a higher correctly resolved LRE rate on both days. As to complexity and accuracy, the results indicate that repeating the same content was beneficial for written target form accuracy and one of the complexity measures (i.e. mean length of T-unit). On the other, in the case of procedural TR, there was a decrease of syntactic complexity and a decrease in the target form accuracy from Day 1 to Day 2. Finally, in terms of learning gains, that is, the pre-posttest results for the target features, procedural TR led to greater gains. In this sense the authors speculate that the more numerous LREs focusing on these forms might have actually marked a difference. Conversely, the lower quality of the procedural TR written production on Day 2 was suggested to be related to the greater cognitive load imposed by engaging with a new topic on that day, which hindered them from handling efficiently linguistic issues.

In EFL settings, the study by Porral Combarros (2017) is one of the pieces of research which has been clearly devoted to the study of these two task conditions. In her MA dissertation study, fourteen adolescent EFL learners (ages 14-16, elementary to intermediate proficiency) carried out the same picture narration task in the same dyads (with homogeneous proficiency) twice. There was a one-week gap between the two performances and learners did not receive any feedback on their writing. Moreover, the interaction of one of the 7 pairs was recorded to analyze their interaction in terms of LREs. The co-written texts were given a holistic mark by an independent teacher and analytically analyzed by the author for syntactic complexity, accuracy and lexical variety. The results showed that while some measures increased (not significantly) from T1 to T2 (number of lexical words, number of clauses, accuracy), others remained the same (accuracy, complexity and lexical density). Regarding the holistic rating, their assessment only reflected a slight improvement from T1 to T2. Finally, as far as the focus of dyad's deliberations is concerned, at T2 learners did not spend so much time on task management and they engaged in writing straightforward, but the author still identified fewer instances of collaboration when the pair repeated the task.

Regarding CW and TR research with YLs, the DG study by Shak (2006) on children's task perceptions is one of the few examples in the literature. The author provided 78 ESL learners in Brunei Darussalam (ages 9-12, mixed proficiency) with a posttask questionnaire after they performed a DG twice on two consecutive days. Although it cannot be considered completely collaborative writing due to the fact that the writing stage was carried out individually, learners did collaborate in the previous stages of the task (such as brainstorming ideas from the listening). The children displayed a more favorable attitude toward DG on the second day, as the task was considered too demanding on the first one. In other words, familiarity had a positive impact on their perceptions on the task.

In the Spanish EFL context, Hidalgo and García Mayo (2019) investigated how TR affected YLs' attention to form while completing three times a picture-prompted story writing in pairs. Forty 11-12-year-old children were to two TR conditions: exact TR (n = 20) and procedural TR (n = 20). The analysis of the LREs showed that in general procedural TR led to more episodes than exact TR, but repeating the task throughout three consecutive weeks did not lead to within-group significant changes between the three CW task performances.

More recently, Hidalgo and Lázaro Ibarrola (2020) set out to explore in more depth the interrelation between CW, TR and YLs by not only looking into LREs but also children's written production quality. In their study, twenty children in a CLIL setting in Spain (age 12, L1 Spanish, A2 level in English) carried out the same collaborative written picture narration task (i.e. exact TR) three times in three consecutive weeks. Their production was assessed by means of a holistic rubric and a range of CAF text-based measures. In addition, they also tallied and classified children's LREs at three different times. Their findings showed that while holistic measures did capture a significant improvement from T1 to T2, T2 to T3 and T1 to T3, CAF measures failed to show any change. On the other hand, the analysis of the LREs revealed that YLs produced progressively fewer episodes throughout the three interventions, and there was a significant decrease from T1 to T3. Yet, most of the children's discussions were devoted to form on all three occasions.

Clearly, more research is needed to determine how procedural TR and CW interact in the case of YLs, and our study aims to shed more light in this respect.

4.2.1.1.3 Task preparation: Pretask FFI

As we have seen, one of the problems of less controlled CW tasks, such as DG, is that despite their focused nature, learners eventually set their own pedagogical goals and concentrate very little (or do not concentrate al all) on the target forms (Calzada & García Mayo, 2020a, 2021a; Eckerth, 2008). One of the possible solutions to this shortcoming has been introducing learners to the target form prior to the collaborative task, for example, by means of Focus on Form Instruction (FFI), such as the provision of a metalinguistic explanation (ME), so that they have a disposition to focus on that feature during task performance. Despite the fact that some TBLT proponents reject any sort of explicit form of teaching because it blurs the main goal of a task (i.e. focusing on meaning rather than on form and obtaining a non-linguistic outcome, Long, 2015), research has demonstrated (mostly, again, from an oral task modality viewpoint) that the risk of this is minimal. Furthermore, ME can "result in more attempts to use the target structure and may sometimes be beneficial for learning" (R. Ellis et al., 2020, p. 59). In this vein, including FFI is felt by other scholars as an excellent complement of TBLT in settings which are characterized by a paucity of exposure to the L2 outside the classroom (Svalberg & Askham, 2020).

As in the case of TR, there is a greater body of research exploring pretask preparation with regards to individual L2 writing than CW. For instance, Shintani et al. (2016) investigated the effect of pretask vs posttask ME on written accuracy. Adult EFL learners in Japan (n = 61, ages 18-20, B1-B2 intermediate proficiency) were assigned to two experimental groups (pretask and posttask ME) and a control group. The target structure was the past counterfactual conditional in English. The writing task involved writing an essay based on some sentences the learners had prepared at home. ME was provided right before or after the writing stage, which was carried out at home. It took the form of a handout explaining the grammar rule. Finally, three individual DG tasks were used as a measure of learners' written accuracy of the target form. The findings from the posttest revealed a significant advantage of the pretask ME group over the control with a large effect size, which was maintained in the delayed posttest.

Another study exploring pretask ME is that by Khezrlou (2019), mentioned in the previous section on TR. In addition to exploring the impact of TR across four individual DG tasks, there were three experimental conditions regarding provision of ME, which took place only at T2: (i) one of the groups received it only before the task, (ii) another before and while carrying out the task, (iii) and finally another one after the task. There was also a control group which did not receive any kind of ME. The ME took the form also of a handout including explanations in the learners' L1 and examples of its use. The pretask and during-task ME group could keep this handout during the tasks while the other conditions could not. The pretask + during ME showed superior writing accuracy than the other ME groups and the control group. This advantage led the author to suggest that the key of explicit instruction may not be the timing but its availability during the writing task.

As far as YLs are concerned, there are doubts regarding the extent to which children can benefit from explicit instruction in a meaningful context owing to the fact that they are "less well equipped cognitively to develop metalinguistic understanding of grammar" (Gorman & Ellis, 2019, p. 62). The study by Gorman and Ellis (2019) did not find strong support for two types of pretask FFI (WCF and ME). In their experiment, they randomly assigned 33 ESL child learners in Kuala Lumpur (ages 9-12, multilingual L1 backgrounds, including English) to one of these two pretask conditions and to a comparison group (n = 11 in each group). The learners carried out four individual DGs (targeting the present perfect) over the course of four weeks and, those in the experimental groups, received pretask focus on form instruction at T2 and T3. The ME took the form of a brief mini-lesson where the teacher elicited the target form in a gapped text and cleared any doubts. The results failed to show any impact of either FFI on the target form accuracy. Yet, the authors suggested that their little impact could have been due to the choice of the target form, as the present perfect could have been beyond YLs' developmental readiness.

Turning now to CW tasks, Storch (2013) explains that pretask grammar lessons have been occasionally used in the literature to foster FonF. One of the earliest examples is that by Swain (1998), where one of the groups was exposed to pretask modelling (a video of the researcher and a teacher performing the task) using metalanguage. In the comparison group, the modelling did not use any metalanguage. The findings revealed that the former group generated more LREs than the latter.

Three more aforementioned studies using collaborative DG have used pretask preparation. Swain and Lapkin (2001) suggested that the lack of differences in the number and type of LREs generated by a collaborative DG and a jigsaw was likely to be due to the grammar mini lesson and task modelling which took place before the task. Kim and McDonough (2011) investigated the use of pre-task teaching and modelling on LREs over three communicative tasks, one of them being DG. Learners who were in the "+ pretask modelling" group produced more LREs and resolved them more correctly. Moreover, they also displayed more collaborative patterns. Contrary to this positive correlation between pretask modelling/instruction and languaging, Leeser (2004), also examining collaborative DG, could not report such benefits in the case of low proficiency SFL learners, who even after being provided that FFI aid still attended more to meaning that to grammar. The present dissertation will try to elucidate how pretask preparation affects attention to form and written accuracy of the target features (3S and POSS) in a collaborative DG task with young low proficiency learners.

4.2.1.2 Individual variables

4.2.1.2.1 Proficiency

In CW tasks, one of the key questions regarding proficiency is related to how similar or different proficiency levels work together. Different studies have investigated this topic in depth. For example, Kowal and Swain (1994) already described the influence of proficiency in their study with collaborative DG in the French immersion context. The pairs were self-selected, and as a result there were homogeneous and heterogenous groupings (although no indication of proficiency level in French is reported). The authors expressed their impressions about these two groupings with regards to the amount of deliberation shown during the task: while in extreme heterogeneous conditions the more proficient learners seemed to take over the leading role and the less proficiency was inhibited, in the more homogeneous dyads the contributions to the task were more balanced. However, since certain degree of heterogeneity also led to collaboration, the authors claimed that this last grouping type was most effective because the learners share some common knowledge but at the same time they can fill each other's gaps.

In another study with collaborative DG, Leeser (2004) also looked into proficiency pairing. In a SFL context, proficiency in Spanish was rated by the learners' instructors as high or low. There were three pairings: High-High (HH), High-Low (HL) and Low-Low (LL). HH generated more LREs, more grammar-focused episodes and more correctly resolved LREs than the rest of the conditions. On the other hand, LL produced the fewest LREs and left a remarkable proportion of them unresolved. What is more, LL's main topic of discussion was lexis. Finally, HL, although they did produce more LREs than LL, the difference did not reach significance. In fact, regarding the mixed proficiency condition, Leeser found that most of the LREs were resolved by the abler learner. These results led the author to question the benefit of the task for low proficiency learners.

In a study which, apart from CW, also involved a reformulated text comparison, Watanabe and Swain (2007) investigated the performance of four *core* intermediate proficiency students (L1 Japanese, adults) who were paired with four higher and four lower proficiency peers. In this case, proficiency was gauged by means of a shorter version of the TOEFL exam. On the first day, the core participants wrote first an essay together with an HP or an LP partner, and afterwards, with a different proficiency peer. The next stage involved noticing the differences of their own text with the reformulated version in each of the grouping conditions. As a posttest, the core learners had to write again the essay by introducing any changes they wished. As a supplement, these participants were also interviewed to tap into their feelings. The recordings of the pair composition and noticing stages in terms of LREs revealed that, notwithstanding the partner's proficiency level, the pairs that showed a collaborative pattern produced more LREs and achieved higher posttest scores.

In a Korean as a Second Language context, Kim and McDonough (2008) compared adult learners' (n = 24, ages 19-25) performance in two experimental DGs, one paired with a same proficiency (intermediate) partner and another with an advanced. In this study, the measure of proficiency was based on an in-house placement test scores. When learners worked with an advanced peer, they significantly produced more lexical and more target-like LREs than when working with an intermediate learner. What is more, learners expressed a preference for working with advanced learners. Nonetheless, the authors noticed that in the heterogeneous condition the lower proficiency member tended to become passive or novice.

Storch and Aldosari (2013) also examined the effect of proficiency on learners' attention to form and, in addition, analyzed the L1 use during task performance. In their study with 36 adult EFL learners (ages not specified), they assigned the participants to one of the three grouping conditions (HH, HL and LL), according to the proficiency shown in their English class activities and previous year high school scores. The learners carried out three collaborative tasks (a jigsaw, a composition and a text-editing). As Leeser (2004), the authors also found that HH generated the highest amount of LREs and LL the lowest. Moreover, the latter were more inclined to discuss lexis than grammar form. These findings study differed from Leeser's results regarding the performance of the mixed proficiency grouping (HL). In this case, they did not always perform better than LL, as the low proficiency member in HL sometimes displayed a passive role which led to very few LREs. In the authors' view, the exceptions to the passive role in HL condition can be related more to affective factors than to proficiency. Interestingly, regarding L1 use, LL tended to rely more on their mother tongue than HL and HH, but still L1 turns accounted for more than half of all turns in all pairing conditions. However, this should not be seen as detrimental. In fact, L1 interaction during collaborative writing tasks has been demonstrated to positively impact lexico-grammatical features of co-constructed texts and facilitate FonF and task management (Zhang, 2019a).

Yet, not all studies looking at the variable of proficiency pairing have focused on its impact on attention to form. One of the exceptions is the study by Shin et al. (2016), where they compared how HH, HL and LL retained ideas from the DG texts, operationalized as ideas units (Carrell, 1985). Thirty-eight ESL adult learners (ages 17-40, age M = 20) were randomly assigned into one of the pairing conditions after they took an in-house placement test. They performed two DG tasks in two consecutive weeks, and the proficiency pairings (homogeneous and heterogeneous) were counterbalanced. However, the results failed to find a significant effect of proficiency pairing on idea unit retrieval.

In the Korean EFL adult context, Choi and Iwashita (2016) delved into the variable of proficiency pairing by specifying the level of proficiency of their participants more objectively. Fourteen Korean L1 learners (age M = 22) had taken the IELTS test no more than a year before the experiment, and according to their scores, there were 8 LP and 6 HP. The *core* participants (as in Watanabe and Swain's study) were the low proficiency learners, who engaged in group collaborative discussions and writing tasks three times, alternating HP-dominant and LP-dominant groups in a counterbalanced design. The task involved watching the beginning of a story and writing in groups a possible ending. The findings revealed that homogenous groups encouraged more positive peer interaction patterns, while in the heterogeneous conditions, LP students were silenced and marginalized.

To conclude this section, a summary of the findings from Niu et al.'s (2018) study is provided. This study was carried out in a Chinese EFL higher education setting. Twenty-four students (ages 19-21) were assigned into HH, HL and LL based on their regular English course scores. The pair writing task consisted in an IELTS graph interpretation task, which they took once. Their performance was video recorded and assessed in terms of LREs and scaffolding strategies. Proficiency pairing was found to affect languaging, but not completely determining it. Hence, HH did not consistently generate more LREs or employ more scaffolding strategies than the other pairing conditions. The authors point out other possible factors which could account for the similarity in those aspects: task nature, interaction patterns and learners' task orientation.

In summary, studies related to pairing proficiency in CW tasks offer mixed findings. In general, there is an agreement that extreme proficiency differences among the members of the same dyad are to be avoided, as the lower proficiency learner can display a passive role and not engage in the task, but at the same time some degree of heterogeneity is advisable (Choi & Iwashita, 2016; Kim & McDonough, 2008; Kowal & Swain, 1994). On the other hand, other studies downplay the role of proficiency, claiming that it does not have any impact on linguistic dimensions other than attention to form (Shin et al., 2015) or suggesting that affective factors play a more relevant role (Niu et al., 2018; Storch & Aldosari, 2013; Watanabe & Swain, 2007). In what follows, we will review studies looking into learner affect and CW. Last but not least, from the review of these studies, we can see that beginner YLs have not been examined in the literature yet. However, adult low-proficiency learners have been claimed not to benefit so much from CW tasks (Leeser, 2004). Therefore, it needs to be clarified to what extent adapting these tasks to the younger population can translate into language gains and L2 writing development.

4.2.1.2.2 Affective factors

The relationships that learners form in CW tasks can influence their performance and, eventually, their learning gains. This has been suggested by several aforementioned studies (Kim & McDonough, 2008; Storch & Aldosari, 2013; Watanabe & Swain, 2007), where the authors examined the effect of other variables, such as proficiency, in addition to their dyadic relationships and demonstrated that the latter could have a stronger impact on focus on form. The reference framework to describe learners' dyadic behavior has been Storch's categorization (2002, 2009a), described above, which established four different patterns in relation to the constructs of equality and mutuality. Researchers who have used this framework have reported a link between collaborative patterns and learning gains. For instance, Storch and Aldosari (2013) found that dyadic relationship played a greater role than proficiency pairing in drawing learners' attention to form.

Although Storch's taxonomy has been proved useful in the CW literature, it is imprecise in nature, as she has admitted (Storch, 2002, 2013). Her recommendation is to carry out "a global analysis, describing the predominant pattern observed in the pair talk in one activity" (Storch, 2013, p. 63), and this is, indeed, the approach undertaken in the present doctoral dissertation. Nevertheless, recent research has attempted to describe the dynamic nature of these relationships, in agreement with the motivational and psychological trend in applied linguistics (Dörnyei & Murphey, 2008; Dörnyei & Otto, 1998), which, although not focusing on pair dynamics, explains the development of L2 learners' relationships throughout the learning process. For instance, Wenxue Chen (2018) analyzed how the patterns of interaction were adjusted by 10 dyads (n = 20, adults, L1 Chinese, B2 upper-intermediate level) during three DG tasks, performed in three consecutive weeks. Moreover, the participants were interviewed after the last DG to tap into their feelings about the task and the partner they worked with. The study confirmed that these adult learners adjusted their patterns of interaction: the general trend was to change non-collaborative patterns present at Time 1 for collaborative or more collaborative ones (e.g. expert/novice). In fact, only one of the pairs remained uncollaborative. What is more, LREs were more frequently generated in collaborative dyads than in non-collaborative ones. The data from the semi-structure interview revealed that learners were well-aware of their relationship during pair interaction, and as familiarity between them grew (participants did not know each other before the experiment), they were more prone to create a collaborative atmosphere.

Regarding YLs, despite some examples of pattern analysis in ESL settings with oral interaction tasks (Azkarai et al., 2019; Oliver, 2002), it is only recently that dyadic relationships among children during CW tasks in FL settings have gained attention. For instance, in an aforementioned study, García Mayo and Imaz Agirre (2019) looked at the patterns of interaction during an oral + written decision-making task comparing three experimental conditions: pairs formed by the researcher (n = 12), pairs formed by the teacher (n = 8) and self-selected pairs (n = 11). The criteria that the researchers used to form pairs was that they should be proficiency-matched, whereas children who self-selected their partners were primarily guided by friendship. The L1 Spanish children were aged 11-12 and had an A2 elementary level in English. The findings revealed that the researcher-selected condition showed significantly more collaborative patterns than the other two groups.

In a study exploring YLs' patterns of interaction during a DG task, Azkarai and Kopinska (2020) examined the relationship between the type of dyadic behaviors of 31 dyads (L1 Spanish, ages 11-12, A2 elementary proficiency) and the depth of engagement of LREs they generated. The authors found that most of their patterns fell under the categories of cooperative (42%) and collaborative (39%), and they reported that the latte produced on average significantly more elaborate LREs than the former (and no differences were found regarding simple LREs).

These studies suggest that both TR and pair selection method can adjust dyadic patterns, and that these interactional behaviors in turn affect the quality of languaging. Nonetheless, there are other individual affective variables which can explain why different relationships are formed. Baralt et al. (2016) found that providing peer feedback and engaging with language during task interaction is related to learners' "attitudes towards task performance as well as towards their partner" (p. 233). Delving into this topic, Sato put forward the construct of "interaction mindset", referring to "a disposition toward the task and/or an interlocutor prior and/or during the interaction" (Sato, 2017, p. 250). In a recent study (Sato et al., 2020) where the construct of "interaction mindset" was validated through Structural Equation Model (SEM), there were five subconstructs (peer interaction, collaboration, form orientation, provision of peer feedback and reception of peer feedback) that went into the final model. In fact, we could argue that these subconstructs could be perfectly compatible with CW tasks.

Although, to the best of our knowledge, this construct has not been applied to CW tasks yet, we will now turn to the body of research that has analyzed learners' attitudes and dispositions towards this mode of writing. According to Ajzen (1988), attitudes refer to "a disposition to react favorably or unfavorably to a situation, an object or an event" (p. 4). Most studies exploring learners' disposition towards CW tasks have been set in ESL contexts. The piece of research conducted by Storch (2005) constitutes one of the first examples. After carrying out a graph interpretation task in pairs, the participants (n = 18, ages 20-42, M 23, mixed L1 backgrounds, high proficiency) were interviewed some days later. On the positive side, learners described the task as entertaining and positive for language learning, while on the negative side, there were a few comments in which they expressed their concerns about their dyad not being collaborative enough. Lin and Maarof (2013) also tapped into adult ESL learners' (n = 30, L1 Chinese) attitudes towards CW but on this occasion participants had to perform in groups of three a summarizing task and, apart from a semi-structured interview, the authors used a 25-item questionnaire. The perception of the task in terms of academic benefits (i.e. not linguistic) was overwhelmingly positive, and the learners also often reported similar views on knowledge co-construction and collaboration. Moreover, the task was considered to increase their self-confidence and to be helpful for grammar accuracy. Yet, the participants acknowledged that sometimes their limited English proficiency prevented them from intervening in certain discussions.

The following ESL studies are related to YLs. Shak (2006) analyzed child learners (n = 78, ages 9-12, mixed proficiency levels) perceptions about the DG task by means of a posttask questionnaire. The setting was a primary school in Brunei Darussalam. However, as explained above, we should bear in mind that the actual writing stage in this study was carried out individually. The participants performed the task on two consecutive days. On Day 1, learners considered that the task too complex and demanding, and consequently their interest and motivation was lower. Nevertheless, as the results from Day 2 showed, gaining familiarity with the task encouraged more positive attitudes. In a follow-up study Shak and Gardner (2008) compared these results from DG with three other form-focused tasks (consciousness raising, grammaring and grammar interpretation), which were also performed twice. The findings revealed that DG was the only task in which child learners displayed better attitudes from Day 1 to Day 2. This result is interpreted from the "Resultative Hypothesis" perspective (K. Johnson, 2008), which predicts that a better grasp of the task will lead to a stronger liking and better performance.

In a foreign language context, Shehadeh (2011) examined adult EFL learners' (n = 18, ages 18-20, L1 Arabic, low-intermediate proficiency) views after a longitudinal CW intervention, which took place over 16 weeks in a university located in the United Arab Emirates. The instrument to collect the data was an open-ended survey. In general, and despite the fact the learners were unfamiliar with this mode of teaching, their opinions were positive and only two participants stated that the instructor should always give a choice to work individually. In a more recent longitudinal study with EFL adult learners as well (n = 2, advanced L2 proficiency, L1 Chinese), Wenting Chen and Yu (2019) gathered the data from two learners who exhibited very different attitudes at the beginning and at the end of the treatment, which took place over 16 weeks. In fact, while one of them shifted from "very favorable" to "very unfavorable", and the opposite was true for his counterpart. The data were collected through multiple sources (video recordings of their CW task performance, surveys, semi-structured interviews, stimulated recall and reflective journals). The findings revealed the dynamic nature of these attitudes, which were mainly influenced by the dyadic relationship they had experienced during the treatment.

Due to the heterogenous proficiency profiles that characterize FL settings, some research has been devoted to studying how the command of the L2 influences CW task perceptions. Gallego (2014) administered a DG task twice to a large sample of SFL learners at two USA universities (n = 497), some of whom were advanced-low (n = 275) and others novice-mid (n = 222), and after the second time, learners were provided with a survey. The questionnaire consisted of 10 items (6 Likert-scale, 3 multiple choice and 1 open-ended), addressing the following topics: (i) the effectiveness of the task, (ii) the perceived skills employed in the task, (iii) the value of metatalk, and (iv) the implementation of metatalk. The results indicated different perceptions according to the learners' Spanish proficiency: the advanced-low students considered the task beneficial for L2 learning, while the novice-mid did not. Furthermore, the latter seemed to be more focused on lexis than grammar, probably due to the difficulties in comprehending the listening), whereas the former could allocate their attentional resources to more dimensions. In this vein, advanced-low learners valued metatalk significantly higher than their counterparts.

Some scholars contend that apart from dyadic relationships and L2 proficiency, group size can also influence attitudes towards CW. For instance, Fernández Dobao and Blum (2013), in a SFL setting (age M = 20, L1 English, intermediate proficiency), compared the views of pairs (n = 28) and small groups of four students (n = 27) on a written picture rearrangement task one week after completing it. The instrument used was a questionnaire which combined Likert-scale and open-ended items, and it addressed several topics such as the perceived usefulness of collaboration (in pairs, small groups or individually) or the impact on the co-written text. The participants showed a clear preference for the grouping mode they had worked in during the experimental session, and more interestingly, they differed in the arguments to favor one grouping or the other. Among the reasons to support pair work, learners stated that dyads were more manageable and they encouraged participation from both learners; on the other hand, arguments supporting small group work included the fact that students could pool a larger amount of resources. The negative opinions about small group work, however, were related to the fact that one of the members tended to remain silent and it was more difficult to organize the task. More strikingly, although these learners enjoyed the CW task, most of them were not aware of any positive influence on linguistic accuracy or L2 learning, as opposed to previous findings (Lin & Maarof, 2013; Shehadeh, 2011; Storch, 2005).

With regards to the influence of group size on perceptions in early EFL settings, Calzada and García Mayo (2020b) explored YLs' (n= 32, ages 11-12, L1 Spanish, elementary proficiency) attitudes towards DG. Some of the participants were assigned to pairs (n = 15) and others to small groups of three children (n = 17). After carrying out the task, learners completed a questionnaire consisting of 12 items, with a combination of Likert-scale, multiple choice and open-ended questions. The survey addressed the following topics: attitudes towards (i) writing, (ii) collaborative work, (iii) collaborative writing, and (iv) DG. Contrary to Fernández Dobao and Blum (2013), these YLs expressed a preference for pair work over small group work, arguing that groups are more difficult to manage and it can be easier for a peer to be silenced. Nevertheless, learners from both conditions reported a high level of comfort working with a researcher assigned partner. As in the case of Gallego (2014), probably due to their low proficiency, participants from both groups failed to acknowledge the opportunities offered by the task for attending to form (instead, they more frequently mentioned the support received from their peer in comprehending content from the listening or writing). Nonetheless, they did value the knowledge sharing and peer feedback. Finally, regardless of the grouping condition, these child learners, in general, displayed a significantly more positive attitude toward writing in L1 Spanish than in L2 English. These differences can, indeed, be due to different writing skills in their L1 and L2, an individual variable that we will explore in the next section.

The most recent piece of research that has looked into the motivation and attitudes of YLs towards collaborative DG is the work by Kopinska and Azkarai (2020). In their study, young Spanish L1 learners of English (n = 64, ages 11-12) performed two collaborative and four individual DG tasks over the course of a school year. Their motivation was tapped by means of a motivation thermometer right before and after each DG enactment. Learners' attitudes were gauged through an attitude questionnaire that was administered at the beginning and end of the experimental procedure (which coincided with the start and end of the school year). The authors reported that certain motivational dimensions (such as anxiety) significantly diminished over time, whereas others (instrumental motivation and motivational strength) significantly increased. Regarding learners' perception of DG in particular, children had a positive attitude towards it both at the beginning and at the end of the school year. Some of the task dimensions these YLs most praised included pair work and the listening stage, while repeating the task (individually and collaboratively) and the writing part were frequently mentioned as the least preferred aspects. Finally, with regards to their motivation, the thermometers showed that right after doing the task their motivation was higher than right before it, and the trend showed a dip every time these child learners performed an individual DG, and an increase when they did so collaboratively. Therefore, the authors found support for YLs' positive disposition towards pair work in a DG task, and also a positive development of their motivation over time.

In conclusion, most research regarding affective factors and CW has concentrated on learner relationships. In general terms, collaborative patterns seem to be a strong predictor of attention to form during task completion. Moreover, collaborative patterns can be encouraged among learners if they are familiar with each other or if there are not extreme proficiency differences among member of the same group or dyad. The attitudes and preconceived ideas about CW that learners bring to the task have also been examined in the literature, which in turn, have been shown to be influenced by L2 proficiency and group size. Despite some attempts to capture the dynamic nature of dyadic patterns during this type of tasks (Wenting Chen & Yu, 2019; Wenxue Chen, 2018), YLs' attitudes towards CW tasks, and more specifically, towards DG task have not been examined in depth. Therefore, we will attempt to examine this topic more in detail in the present dissertation.

4.2.1.2.3 L1 writing skills

As we have concluded from the previous review, both L2 proficiency and affective factors influence collaborative L2 writing. What is not so clear is whether writing expertise in the L1 affects the texts produced in the L2 and the writing process (Manchón & Williams, 2016). This question has been tackled from an L2 writing perspective, one of the three main domains that CW research draws on, but also from the emergent field of multilingualism.

The departing point is the theory of multicompetence, developed by Cook (1992), which posits that the knowledge of a bilingual or a multilingual is not the sum of the knowledges of two monolinguals, but it rather constitutes a compound state of competence. This theory is intertwined with the concept of cross-linguistic influence. Studies looking into cross-linguistic influence took off in the Canadian immersion setting (Genesee, 1987), where the writing of bilinguals in French and English was assessed. Although cross-linguistic influence was considered negative for a long time, its positive impact, as well as the different directionality in which it may occur, that is, both from the L1 to the second/subsequent language(s) (Ln) and vice versa (Jarvis & Pavlenko, 2008; Ringbom, 2007), has in more recent times begun to be acknowledged. As Schoonen et al (2011) explain, although certain knowledge resources can be language-specific (such as vocabulary and grammatical knowledge), "higher order, less language-specific skills, such as metacognitive knowledge and skills concerning effective writing strategies, could be useful across languages" (p. 36).

The scope of topics examined by studies exploring the relationship of L1 and L2/Ln writing covers the product and process of writing (Rinnert & Kobayashi, 2016). The different objects of analysis are divided between the *repertoire knowledge of the writer* (including topic knowledge, L1/L2/Ln writing knowledge and genre knowledge) and the *social context* (all the factors that influence writer's decisions, such as their perceptions, attitudes, language proficiency, recency and other situational factors, such as the task). An example of a study in this domain would be to analyze, from a contrastive rhetoric perspective, how genre differences in the L1 and L2/Ln influence learners' text discourse characteristics. Hiroshe (2003), for instance, studied how L1 Japanese writers tackled essay argumentation in L2 English, two languages where this genre conventions differ notably.

Another example, in this case from the Basque EFL setting, can be found in Cenoz and Gorter (2011). In their study, they looked at the Spanish, Basque and English written production of 165 secondary school students (age M = 14.6; L1 Spanish, L1 Basque and bilinguals). The participants completed a picture description task on three different days for each of the languages, and they were required to write 250 words in all of them. The production was later analyzed by means of an analytic rubric (Jacobs et al., 1981).

Additionally, the authors examined the lexical and phrasal borrowings in all directions. The correlation analysis between English and Spanish and Basque did not render high values. However, they demonstrated that multidirectional transfer occurred in each of the three languages, which suggests that learners established "soft boundaries" (p. 366) between them. In many cases, those strategies were used to compensate for the lack of knowledge, but in others the justification was more related to affective factors.

There are other studies which have taken a more SLA-oriented approach to examine the L1-L2 writing connection. Within this orientation, some research has investigated, for example, the cognitive processes that take place during both writings by means of think aloud protocols and posttask interviews. Manchón et al. (2009) based their investigation on Flower and Hayes' (1981) L1 writing model, which posits a cyclical interplay of three main processes known as planning, formulation and revision. They compared how different age and proficiency EFL learners allocated task time to each of those processes in the L1 and in the L2. There were 21 participants involved (n = 7 secondary school with a B1 level, n = 7 higher education with a B2 level and n = 7 higher education with a C1 level). The writing tasks required learners to compose argumentative and expository texts. The results showed that, in contrast to formulation, there were great differences in planning between the two languages, and that these differences were constrained by the participants' L2 proficiency. B1 and B2 proficiency students planned more in the L1 than in the L2 writing task, while the C1 students did the opposite.

Finally, Manchón et al. (2009) were also interested in determining to what extent the problem-solving activity differed according to the language in which participants were required to write. The findings indicated that L2 writing involved more problem-solving due to the numerous problems that learners encountered (twice as many as in the L1). Furthermore, the classification of the problem-solving deliberations demonstrated that these did not just occur as a result of L2 knowledge gap, but that there were, indeed, examples of "compensatory" and "upgrading" strategy use. The former refer to instances when learners make up for knowledge deficit, whereas the latter imply improving linguistic choices. Upgrading strategies were more frequently used by C1 level students, who did not experience a performance loss from the L1 to the

L2 thanks to their higher proficiency. The B2 level participants, in contrast, had not reached the threshold that enables writing skill transfer from the L1 to the L2, hence showing a wider use of compensatory strategies.

Moreover, the type of linguistic questions learners dealt with in L2 writing was influenced by their L2 proficiency, as higher-level concerns (for example, related to discourse) were only attended to by those L2 learners with a higher command. In general, the use of the L1 while composing in the L2 was wide; yet, the L1 was used differently in each proficiency group. The lower proficiency learners used it as a compensatory resource, whereas the more advanced learners employed their mother tongue for processing-planning, organization and rhetorical and discourse problems.

Another group of studies with a clear SLA impetus has tried to quantify, usually by means of Structural Equation Modelling (SEM), the extent to which L1 writing expertise explains L2 writing variance, compared to a set of other individual variables (such as affective factors). Schoonen et al. (2011) analyzed longitudinally (in three testing points over 2 years) the development of L1 and EFL writing of 400 secondary school Dutch learners. Apart from the written production data in each of the languages, the authors gathered data about the participants' metacognitive and linguistic knowledge and linguistic fluency (i.e. speed of processing of lexical and grammatical information). At the beginning of the study (ages 13-14), L1 writing strongly correlated with L2 writing, and in fact, it accounted for variance which could not be explained by any of the English language subskills. Yet, at Time 2 and 3 the prediction model did not improve with the inclusion of L1 writing skills. The authors clarify that this does not mean that there was no relationship, but rather that it might have been stable for the course of two years.

Pae (2008, 2018) also found support for the association between L1 and L2 writing. In the first study, the author examined the causal relationship between L2 writing quality and five independent variables (including L1 writing ability, attitudes to L2 writing, cognitive knowledge of L2 writing and free reading and writing practice outside school). Sixty-six Korean adult EFL learners (age M = 22) wrote an argumentative essay in English and Korean, and took a questionnaire in order to gauge the rest of independent variables. The findings revealed that L1 writing ability directly affected L2 writing quality, whereas attitudes towards writing were indirectly related to L2 writing by means of L2 writing practices outside school. In other words, only the L2 writing scores of those learners who practiced L2 writing in their free time were positively influenced by their attitudes towards writing.

In the second study (Pae, 2018), this relationship was examined together with the moderating effects of task complexity and language proficiency. Three hundred and five Korean EFL learners (ages 16-17) were required to write one expository (- complex) and one argumentative (+ complex) in each of the languages. Regardless of task type, the author found moderate correlation between L1 and L2 writing, and proficiency proved to be an important moderator, as the coefficient was higher for the high proficiency group. This finding was considered by the author to be evidence of the Linguistic Threshold Hypothesis (LTH) (Alderson & Urquhart, 1984; Clarke, 1980; Yamashita, 2001), a development of Cummins' theory of Linguistic Interdependence (Cummins, 1979). According to the LTH, the magnitude of the correlation between L1 and L2 skills will depend on the current L2 proficiency level, and for successful L1-L2 transfer to occur, a certain threshold needs to be crossed.

Although the aforementioned studies used subjective measures of writing (such as holistic and analytic rubrics) to assess performance in the L1 and the L2, Ströbel et al. (2020) employed 12 text-based complexity measures. Eighty adult EFL learners in Germany (age M = 23.92) with a B2 level in English contributed to the corpus of the study by providing a sample of an academic writing in the L1 and the L2 (hence, the study took a natural language production approach). The authors found support for Cummins' theory of Linguistic Interdependence, as the statistical analyses revealed a positive relationship between L1 and L2 writing for eleven of the twelve measures. Furthermore, this correlation continued to be so even after controlling for individual confounding variables, such as age and time spent in an English-speaking country. Thanks to the use of the same text-based measures in the L1 and in the L2, the authors confidently claimed that the relationship was independent of the type of linguistic knowledge.

Last but not least, a few studies have investigated the L1 writing experience and perceptions that learners bring to L2 writing. For instance, Silva (1992) demonstrated that adult ESL learners were aware of a large amount of differences between L1 and L2 writing, especially with regards to the dimensions of planning, writing process, grammar and vocabulary used in each of them. In a more recent study, Saeli and Cheng (2019), in the Iranian adult EFL context (n = 15; L1 Farsi, ages 20-30), showed that previous L1 writing experiences shaped learners' perceptions of L2 writing. More specifically, the lack of emphasis on writing in the L1 curriculum and its old-fashioned instruction (heavily relying on literature and grammar) was identified as the main reason for students' failing to appreciate the value of writing in English (as opposed to other skills, such as speaking). In the Primary EFL context, Calzada and García Mayo (2020b) attempted to tap into YLs' perception of L1 and L2 writing in an exploratory study (n = 32, ages 11-12, L1 Spanish). The results showed a significantly more positive disposition towards the former than the latter.

After this brief review on the L1-L2 writing connection, it is considered that there are enough reasons that justify the need to explore the influence of L1 writing skills as a moderator of L2 writing achievement both when working individually and in collaboration. Once again, the lack of studies targeting YLs in EFL settings indicates the need to widen the research agenda in this domain as well and, hence, in the present dissertation, we intend to investigate this topic exploratorily.

4.2.2 Measuring writing: The CAF framework

Just as in L2 writing research (Polio & Friedman, 2016) and in line with task-based performance research (R. Ellis et al., 2020), CW studies have embraced the *complexity, accuracy, fluency* (CAF) framework to analyze learners' written production (Zhang & Plonsky, 2020). This triadic concept has been recently supplemented with the area of lexis (L), as a separate construct from syntactic complexity (Bulté & Housen, 2012). Furthermore, in order to attend to the communicative requirements imposed by tasks, some researchers have suggested adding a fifth component to the list, known as Functional Adequacy (FA) (Kuiken & Vedder, 2018; Pallotti, 2009). Therefore, although sometimes the acronym to refer to this framework has been extended to CALF or CALF-FA, in the present dissertation, for the sake of consistency, we will employ CAF, as it still the most common in the literature.

The *complexity, accuracy* and *fluency* framework has been employed in L2 writing since the inception of the field (Wolfe-Quintero et al., 1998), as a means to quantify L2 development. R. Ellis et al. (2020) summarize the main justifications for this approach:

(i) Several factor analyses have demonstrated that the constructs are independent from each other (Skehan & Foster, 1997; Tavakoli & Skehan, 2005). This means that high scores in one of the subdomains may not translate into equal performance in the other dimensions. Likewise, a certain learner internal variable (e.g. proficiency) may not equally affect them.

(ii) Some research (Skehan, 1998, 2004) suggests the existence of a developmental sequence which affects all components of CAF. First, learners strive to gain complexity (both syntactic and lexical), and subsequently, as their interlanguage stabilizes and grows, they have a better command of the L2 which allows them to decrease errors (i.e. increase their accuracy) and, finally, to speed up their production (i.e. to increase their fluency). However, the multifaceted nature of CAF is nowadays widely acknowledged, as well as the fact that it can show strong individual variation (Gunnarsson, 2012). In fact, some measures cannot be expected to grow in a linear fashion (Larsen-Freeman, 2006).

(iii) In addition to the cognitive or task-related variables influencing the development of CAF in diverse ways, some research has shown that these dimensions vary depending on learners' own stylistic preferences (Pallotti, 2009; Skehan & Shum, 2017).

The findings from TBLT studies using the CAF framework have usually been interpreted from a psycholinguistic perspective, more specifically, in terms of two competing accounts: Robinson's (2001) Cognition Hypothesis and Skehan's (1998) Limited Attentional Capacity. The differences between these two hypotheses lie in how the different constructs will interact under certain task conditions (e.g. task complexity) given the attentional or cognitive resources that we have.

Skehan predicts that, since our attentional capacities are limited, the more complex a task, the more likely trade-off effects will be (i.e. there will be competition between CAF dimensions), especially between accuracy and complexity. Conversely, Robinson puts less emphasis on the learner's attentional capacity and contends that parallel increases in complexity and accuracy can occur provided that certain task conditions are met. Moreover, Robinson hypothesizes different CAF outcomes in relation to, on the one hand, the interference or competition between resource pools (divided into three categories: processing stages, modality and codes of processing) and, on the other hand, the resource-dispersing and resource-directing elements involved in a task. An example of a resource dispersing task condition would be requiring learners to perform more than one task simultaneously, which can lead to attend to non L2-specific areas. On the contrary, a resource-directing feature could be to require learners to refer to many elements instead of few, which consequently would encourage them to concentrate on specific L2 constructions. Likewise, similar predictions based on these two accounts have been made taking TR as a variable, but there is more consensus that CAF measures will increase if learners encounter a task more than once (Michel, 2017)

Although these models were conceived for short-term oral tasks, and hence, many of the task condition variables contemplated in them (such as planning time or TR) ought to be rethought for the writing modality (Manchón, 2014b), empirical studies in the domain of individual L2 writing provide conflicting results and there is no strong and consistent evidence to support one of the two models (Kormos, 2011; Ong & Zhang, 2010, 2013). Regarding CW, product-oriented studies (Wigglesworth & Storch, 2009), despite employing CAF measures, have not interpreted the results in terms of these cognitive models (with the exception, perhaps, of Kuiken and Vedder, 2002a). As mentioned earlier, CW studies have been primarily entertained from an SCT optic, which posits that task performance will always depend on each individual's interpretation of that task. Therefore, these studies have not been so interested in pointing at the source of the variability in written performance, but in the actual different understandings of tasks.

Despite the popularity of CAF measures in L2 writing research and the large amount of studies that have demonstrated their validity and usefulness, this framework (which does not constitute a theory *per se,* Pallotti, 2009) is not exempt from criticism. One of them is related to the use of oral measures to describe written production, for example in the case of syntactic complexity (Biber et al., 2011). In fact, the majority of research that has employed these measures has examined oral research only (R. Ellis & Barkhuizen, 2005). Yet, probably the most important shortcoming of CAF is the sheer variety of measures employed by researchers; for instance, as reported in Zhang and Plonsky (2020), in the field of CW, 15 measures of syntactic complexity were used in 17 studies. This lack of measurement consistency across studies makes it especially difficult to establish comparisons and draw general conclusions about the development of CAF in this modality, and it is hence not infrequent to find conflicting results in the literature (Skehan & Foster, 2012). Nonetheless, employing different CAF metrics is justified by the need to cater for ecological validity. As Norris and Ortega (2009) state:

> [R]esearchers must engage in a much more organic practice in order to achieve understanding of CAF as conditioned by the realities of learning contexts […] it means that measurements will also need to provide learner-, task-, and L2 form-sensitive accounts of the local SLA ecology, given the ways in which these factors moderate the observation we might be making about CAF (p. 574)

This view is also shared by Michel (2017), who argues that the choice of the measurement will very much depend on the L2 data involved. Therefore, she suggests that apart from using key reference metrics in the field (such as errors per 100 words, in the case of accuracy), we should add measures which are specific for the current study and the research questions to be addressed. In the present dissertation, for instance, general accuracy is examined, but also target form production (3S and POSS) as well as content accuracy, in terms of the amount of Idea Units (Carrell, 1985) retrieved from the original DG texts, are specifically examined. In what follows, we will briefly define each of the CAF constructs and explain the most used measures:

Syntactic (or structural) complexity

Syntactic or structural complexity has been discussed extensively in L2 writing and TBLT research. As Ellis et al. (2020) explain, there are two approaches to the study of this construct: one has to do with the range of structures employed by the learner in a writing task, and the other is related to the calculation of a measure of subordination. The former approach hypothesizes that the more structurally varied a text is, the more complex it will be, whereas the latter offers a more reduced vision, by considering subordination as the primary mechanism which learners resort to for conveying more condensed information. Ideally, studies should include both type of structural complexity measures, as subordination is not expected to appear in lower level YLs' production (Torras et al., 2006), and it may not discriminate well in advanced levels (Norris & Ortega, 2009). An example of a recent study in which both kinds were used is Ahmadian and Mansouri (2020), where the co-constructed DG texts were assessed in terms of the ratio of clauses to T-units (syntactic complexity) and also according to the number of different grammatical verb forms, focusing on tense, modality, and voice (syntactic variety). In the case of young EFL learners, recent corpus-driven research has suggested that noun phrase complexity can be a good discriminator of syntactic complexity (Díez-Bedmar & Pérez-Paredes, 2020). Finally, it should be remembered that although there is a tendency to interpret more complexity as better language use (e.g. more subordination indicating a stronger command of the L2), this view is considered too simplistic (Pallotti, 2009), because, on the one hand, it can depend on personal stylistic preferences, and on the other, given the non-linear and dynamic development of CAF, more subordination could be in detriment of other areas.

Lexical complexity

Lexical complexity has been subdivided into three categories (Jarvis, 2013), namely, lexical diversity, lexical sophistication and lexical density. The first has been probably most studied and operationalized in the literature. It refers to the "range of different words used in a piece of writing" (Polio & Friedman, 2016, p. 108), and it has been usually calculated based on type-token ratios. However, as these ratios have been shown to be influenced by text length (Skehan & Foster, 2012), many researchers have opted for one of the corrections, such as Guiraud's Index (GI) (Guiraud, 1959), which will be further explained in the "Coding" section. Another popular measure is D, which is calculated through a complex statistic based on type-token ratios. Meara and Miralpeix (2017) created a program (D_Tools v20) to compute this automatically based on previous work on *D* (Malvern et al., 2004). Yet, the texts that are entered in the software need to be of a minimum length, as the program generates 100 samples of 35 words.

Lexical sophistication, conversely, aims to gauge the construct of "lexical richness", and it analyzes the proportion of infrequent words in the written production. This is calculated, for instance, by counting the words which are beyond the 1000 most common lexical items in the language or text genre. Meara and Miralpeix (2017) also developed a software (P_Lex v3.00) to automatically do this with English texts. However, texts need to have a considerable length, as the program subdivides the text into ten segments. Skehan and Shum (2017) argue that whereas lexical diversity can clearly distinguish between native and non-native speakers, lexical sophistication does not. In fact, the former is considered an ability of the learner, whilst the latter appears to be more task-related.

Finally, lexical density, the less studied construct, refers to the amount of information contained in a text, which is understood as the ratio of lexical words per function words (Michel, 2017). However, this construct is generally less useful with low-proficiency students as their writing resembles a telegraphic style and at this stage even function words might already be indicative of their lexical knowledge.

Accuracy

Accuracy can be defined as the extent to which an L2 learners' performance conforms to the target language (Housen et al., 2012; Pallotti, 2009). Although this construct may give the impression of being the most transparent of all, some researchers have pointed out that it can lead to issues regarding its validity and reliability (Polio & Shea, 2014). Firstly, there is debate about what constitutes an error, as this question deals with prescriptivists views on language which are very much tied to the use of a certain language variety. Consequently, two or more raters may not always agree on what is accurate in the learners' written production (Kuiken & Vedder, 2014). Secondly, there is some consensus that not all errors are equally severe, for instance, omitting an article may not hamper communication to the same extent as omitting a comma or a full stop. Typical accuracy measures, such as proportion of error-free clauses or errors per 100 words (Ellis et al., 2020) treat all errors the same. Therefore, some researchers have proposed the concept of *error gravity*, which aims at controlling for errors' hindering impact on communication (Foster & Wigglesworth, 2016). Error gravity assigns a score to each clause in the text based on its accuracy. However, this task is not exempt from subjectivity, and it should be acknowledged that any sort of accuracy analysis will always involve occasionally "some degree of personal judgment" (p. 112). Yet, unlike complexity measures, which do not correlate highly with each other, it has been shown that it is very likely to come to the same conclusions about accuracy using any of the proposed measures (Skehan, 2018).

Finally, regarding YLs' production, accuracy can even become more complex. In fact, some of the most commonly used measures (error-free units) can be problematic given the short length of the compositions and the many orthographic mistakes, which may be the product of a developing literacy (Tejada-Sánchez & Pérez-Vidal, 2018). Hence, the literature suggest the use of more fine-grained measures, (i.e. specifying the structural nature of the errors, e.g. grammar, lexical and mechanical errors, as in Villarreal & Gil-Sarratea, 2019) or taking into account only those typical for their age and level of proficiency, a criterion which usually involves consulting the children's teachers in advance.

Fluency

Although the construct of fluency has been explored to greater depths in the case of oral production (Michel, 2017), in the written mode it has involved rating metrics (e.g. number of words produced divided by the total task time) or length (e.g. number of words per sentence, clause or T-unit). More advanced ways of measuring computer-based written fluency can also capture the writing process itself (pausing, revising, reformulating and the cognitive processes underlying these behaviors), for example, by means of keystroke logging (Révész et al., 2017). Finally, the scarce body of literature addressing task-based written performance by EFL children has used both raw frequency (number of words, sentences and clauses), length (words per sentences and per clauses) and rate measures (proportion of words over task time) to gauge fluency (Tejada-Sánchez & Pérez-Vidal, 2018; Torras et al., 2006).

Text quality measures

Along with CALF-FA measures, CW research has also examined text quality by means of analytic and holistic rubrics[11], although the former have been used to a greater extent than the latter (Zhang & Plonsky, 2020). Analytic rubrics usually measure

[11] Assessment based on rubrics has also been referred to in the literature as "qualitative analysis", as opposed to CAF measures, which have been termed "quantitative analysis" (Storch, 2005). However, we believe that this terminology can be misleading, given that rubrics involve a numeric score which makes them quantitative in nature.

different dimensions separately, whereas holistic rubrics provide a general score for the text (Weigle, 2012).

One of the most popular analytic rubrics in the field of L2 writing is the one developed by Jacobs et al. (1981) for ESL writing profiling, which consists of five categories (organization, vocabulary, language and mechanics). Each of the subscales carries a different scoring weight: with content having the highest weight (30% of the total score). Moderate weights are given to language use, organization and vocabulary (25%, 20% and 20% of the total mark, respectively), while mechanic receives the lowest (only 5% of the total mark). Shehadeh (2011), for example, employed this scale in his longitudinal study of CW with adult EFL learners in the United Arab Emirates. However, some authors have considered these proportions unsuitable for the EFL context (Ghanbari et al., 2012), where mechanics, for instance, can be a primary concern at beginner levels of the language. Hence, there have been different adaptations of Jacobs et al.'s rubric, such as the one proposed by Villarreal and Gil Sarratea (2019) for assessing EFL teenagers' CW production, where each of the five subconstructs is weighted equally on a four point scale. Other CW studies developed an ad-hoc rubric to answer the researchers' particular questions, as in the case of McDonough and García Fuentes (2015). These authors assessed adult Thai EFL university learners' collaboratively written paragraphs based on an analytic scale addressing three categories (content, organization and language), rated on a 4-point scale (poor, needs improvement, satisfactory and good). In fact, the authors acknowledged that the rubric employed was adjusted based on the university instructors' feedback prior to the experiment. In a later study in the same context, McDonough et al. (2018) used the same three dimensional rubric, but this time on a 1-10 scale, along with descriptors for each of the bands.

Conversely, examples of holistic rubrics can be found in Storch (2005), where she assessed learners' text writing in terms of structure and task fulfilment using a 1-5 scale and each score band was accompanied with a brief description. Finally, examples of the use of rubrics to assess YLs' writing are still scarce. For instance, Calzada and García Mayo (2021c), in order to supplement the information gathered from the text-based measures, resorted to the evaluation rubric employed by the regional government to assess EFL narrative writing in the last year of primary school (Department of Education, 2020).

Text quality measures based on rubrics, whichever the type, have the advantage of being more ecologically valid, inasmuch they resemble the most usual way to assess learners' writing in the L2 classroom. Indeed, the CEFR (Council of Europe, 2001, 2018) contains a myriad of rubrics to assess learners' linguistic competences, including writing, and these tend to be employed by L2 instructors. However, their validity can also be compromised when the raters are not able to distinguish, for instance, accuracy from other dimensions such as length and content (Polio & Shea, 2014). Therefore, it is always necessary to precede their experimental use with a piloting stage, which according to Zhang and Plonsky's review, is not typical in CW studies, and to analyze intercoder reliability to ensure that when others use our rubric, they obtain similar scores.

To summarize, in Part I (Literature review), the main topics surrounding the current dissertation have been reviewed. First, in Chapter 1, the main characteristics of child L2 learning have been explained and the historical and social background of current early EFL policies have been provided, focusing on the specific case of Spain and the BAC. Then, in Chapter 2, the TBLT framework has been examined and the theoretical justification from the interactionist, sociocultural and psychological paradigms for its implementation in the L2 classroom has been offered. The next section, Chapter 3, has been devoted to the application of TBLT with YLs, and the most relevant empirical research to date in this area has been outlined. Finally, in Chapter 4, we have argued for the use of CW tasks (including the benefits of collaboration, the writing-to-learn approach or the CW task typology), and we have thoroughly analyzed the studies that have previously used the DG task. This chapter has been concluded with the review of the variables that influence CW tasks and with the justification for assessing written production from the CAF framework. In the next section of this dissertation (Part II), the empirical study is presented.

PART II: The present study

CHAPTER 5 THE STUDY

5.1 Motivation for our study

In Part I, several research gaps were identified. First, given the popularity of early EFL programs, it is striking that not more studies have been devoted to YLs (Enever, 2018; García Mayo, 2017). In fact, different authors (Muñoz, 2008; Murphy, 2014; Tellier & Roehr-Brackin, 2017) contend that generalizing findings from adult EFL and early ESL settings is inappropriate, as there are certain features specific of early EFL, such as minimal exposure to the FL outside the classroom, a later starting age than in ESL settings, a lower memory and attentional capacity than adult learners, a complete literacy in the L1 by the end of primary education allowing for crosslinguistic influence and metalinguistic reflection, to mention only a few.

Secondly, in the era of the Communicative Language Teaching, TBLT is still considered an innovation because it emphasizes “learning-by-doing” and real-life tasks (with an outcome that transcends the purely linguistic one) as opposed to other methodologies or frameworks (R. Ellis, 2018). In this vein, tasks have been used in SLA both as a window allowing researchers to observe L2 learners’ linguistic performance and knowledge under a myriad of task-related and individual factors, and, on the other hand, as a means to inform (and improve) L2 instruction. However, empirical TBLT studies have mainly focused on adult ESL/EFL (Wen & Ahmadian, 2019) and to a much a lesser extent on child ESL learners (Azkarai et al., 2019; Oliver, 2002). As it can be inferred from the review of TBLT research, the oral mode has traditionally dominated in this field, as Manchón and Williams (2016) argue, probably influenced by the desire to obtain spontaneous learner production. Yet, the benefits of L2 writing are supported by a number of psycholinguistic arguments, including the possibility of turning explicit knowledge into either proceduralized explicit knowledge (DeKeyser, 2007) or into implicit knowledge (N. Ellis, 2011), thanks to a larger amount of time that enables continuous reflection and revision of the production (Williams, 2012; Zalbidea, 2021). Taken together, two main gaps can be found: a lack of TBLT research devoted to YLs in EFL settings, which has only recently started to be explored (Butler, 2017; García Mayo, 2018b; García Mayo, 2021c; García Mayo & Gutiérrez Mangado, 2020; Newton & Le Diem Bui, 2020; Pinter, 2015; Shintani, 2014), and the need to strengthen the L2 writing-to-learn and TBLT connection (Byrnes & Manchón, 2014).

Thirdly, although TBLT research has been informed by multiple paradigms, studies on CW have primarily drawn on SCT (Storch, 2019a). Influenced by the early work of Swain (Swain & Lapkin, 1998 *et passim*), interest in the collaborative dialogue resulting from these tasks is based on the assumption that it constitutes "language learning in progress" (Swain & Lapkin, 1998, p. 321), and not just facilitative of the cognitive mechanisms which allow to further L2 knowledge, as scholars from the cognitive strand hold (Gass & Mackey, 2006; Long, 1988, 1996). The most widely used task in L2 CW research has been the DG task (Storch, 2016), but as concluded from our systematic review, it has been mainly investigated with adults and teenagers, mirroring the general trend in studies on TBLT (Plonsky & Kim, 2016) and CW research (Zhang & Plonsky, 2020). Nevertheless, the use of DG has been shown successful in terms of drawing YLs' attention to form over meaning and generating LREs (Calzada & García Mayo, 2020a, 2021a), although difficulties have also been reported when it comes to focusing children's attention to the target forms or furthering L2 grammar knowledge.

In addition, CW tasks have been claimed to be influenced by task-related and individual factors. However, most studies have looked into the impact of task type (meaning focused vs form-focused) (Alegría de la Colina & García Mayo, 2007; García Mayo, 2002b; Storch, 2001) and learner proficiency (Choi & Iwashita, 2016; Kim & McDonough, 2008; Kowal & Swain, 1994; Leeser, 2004; Niu et al., 2018; Shin et al., 2016; Storch & Aldosari, 2013; Watanabe & Swain, 2007), leaving other variables underexplored. Yet, evidence from research on oral tasks suggests that adult and child task performance can be influenced by TR (Ahmadian et al., 2017; Bygate, 2009; García Mayo et al., 2017; Lázaro Ibarrola & Hidalgo, 2017) as well as affective factors (Azkarai et al., 2019; Baralt et al., 2016; Oliver, 2002). Moreover, research on individual DG with children and adults has explored how the inclusion of FFI can impact learners' attention to form (Gorman & Ellis, 2019; Khezrlou, 2019).

Finally, although the main goal of CW has been to describe the task process, either by quantifying the amount of FonF (i.e. LREs), or by assessing the quality of the written production, increasingly more studies are looking into how CW can create changes in learners' L2 knowledge, including L2 writing expertise (Qiu & Lee, 2020). In other words, a task-as-treatment or experimental approach is now being encouraged (Elabdali, 2021). In order to do so, researchers have employed the CAF framework, and a few scholars have tried to apply these dimensions to research with YLs' (Calzada & García Mayo, 2021c; Hidalgo & Lázaro Ibarrola, 2020; Torras et al., 2006).

The present study intends to shed some light on the aforementioned research gaps by analyzing YLs' (ages 11-12) performance in two DG tasks. These tasks were conceived to draw their attention to two problematic forms: the 3rd person singular present tense marker *-s* (3S) and the third person singular possessive determiners his/her (POSS). In order to compare the benefits of interaction, a comparison group who carried out the tasks individually was included. Likewise, there was another experimental group who, in addition to experiencing the collaborative DG, received pretask FFI on the two linguistic features, so as to facilitate their noticing during task performance. Furthermore, before and after the task learners carried out an L2 writing task to measure any changes in their individual L2 writing production. In order to control for the moderating effect of L1 writing expertise, they also completed an L1 writing test prior to the treatment. In addition, we aim to obtain a better understanding of YLs' affective variables surrounding the L2 writing, collaborative work and the DG task. In order to so, before the DG intervention, the participants completed a questionnaire on their attitudes towards writing and collaborative work. Finally, after the pedagogical intervention, their views on the dictogloss task were tapped into by means of a questionnaire and focus groups.

In what follows, the research questions (RQs) that guided our investigation are presented and the description of the material and procedures used is provided. To conclude the section, the data codification and analysis is explained.

5.2 Research questions and hypotheses

After the review of the literature and before conducting the study, the focus of the investigation was narrowed down by defining the RQs. These were classified into three broad topics:

- RQ1: Dictogloss, Focus-on-Form (FonF) and the impact of task related-variables
  a) What is the focus of YLs' written languaging (wLREs) and oral language-related episodes (LREs)?
  b) What is the LRE outcome, depth of engagement and Previously Known Language (PKL) use?
  c) What is the impact of procedural task repetition (TR) and pretask focus on form instruction (FFI) on YLs' focus on form?
  d) Do YLs focus on the two target forms (3S and POSS)?
- RQ2: Collaborative work
  a) What are YLs' patterns of interaction during collaborative dictogloss? Are they influenced by TR?
  b) To what extent do collaboration, TR and FFI influence the quality of the dictogloss written product?
  c) Does dictogloss have an impact on subsequent individual writing?
- RQ3: Individual factors
  a) What are YLs' attitudes to writing and collaborative work?
  b) What are YLs' insights about the dictogloss task?
  c) What is the impact of YLs' L1 writing skills and attitudes towards writing on L2 writing?

Taking into consideration the previous findings in relation to each of the RQs, these hypotheses were formulated:

RQ1a: Studies investigating low-proficiency EFL YLs and collaborative DG (Calzada & García Mayo, 2020a, 2021a) have shown that children are generally able to focus on form in a communicative task. Furthermore, in line with the results of this research, we hypothesize that learners will focus more on formal aspects of writing (i.e. grammar and mechanics) than on lexis.

With regards to individual written languaging (wLREs), although it has been much less extensively studied in the field of L2 writing, research by Ishikawa (2018) and Ishikawa and Révész (2020) has shown that providing adult EFL learner with opportunities to reflect in writing about their language use after an individual DG task can promote the explicit knowledge about the target form (in their studies, the present counterfactual conditional), more specifically, regarding production accuracy scores. However, the comparison of the advantages of wLREs and oral LREs with adult L2 participants (Watanabe, 2019) has found that, even though WL may have certain positive impact on productive measures of our target forms, the benefits generated by peer work and collaborative dialogue are greater. Thus, it is expected that oral LREs will display greater metalinguistic reflection than wLREs in the case of YLs.

RQ1b: We hypothesize that child learners will be able to correctly resolve most of their metalinguistic discussions and that the proportion of unresolved LREs will be low (Calzada & García Mayo, 2020a, 2021a). Regarding the depth of engagement, Azkarai and Kopinska (2020), using Storch's (2008) classification of engagement, found that child EFL learners completing a dictogloss task mostly produced Elaborate LREs, followed by Limited and Limited + Limited episodes. Yet, Calzada and García Mayo (2020a) reported that the engagement of LREs devoted to target forms in particular (3S and articles) varied depending on how familiar learners were with the target form prior to the treatment. As a result of having received more previous instruction on the 3S than on articles, learners produced more elaborate LREs in the former and more limited ones in the latter. However, thanks to the careful selection of target forms in the present dissertation, it is foreseen that learners will produce more elaborate LREs than limited ones in both 3S and POSS.

Finally, regarding L1 and L2 use, although to our knowledge no previous study has analyzed this variable concerning collaborative dictogloss, recent research with Spanish YLs of English performing oral and oral + writing tasks (Martínez Adrián & Arratibel Irazusta, 2020) has demonstrated that children do rely on their mother tongue(s), especially in tasks with a writing component. Since most PKL has been shown to be related for metacognitive functions and vocabulary searches, being LREs the object of our analysis, we anticipate a greater use of PKL than L2.

RQ1c: Regarding these two task implementation variables, different hypotheses are formulated. Firstly, TR has been shown to produce mixed findings with regards to metalinguistic engagement in CW tasks. On the one hand, adolescent EFL learners appear to spend less time in task management, but this extra time does not translate in more LREs (Porral Combarros, 2017). On the other hand, in the case of young EFL learners, LREs were unaffected by TR (Hidalgo & García Mayo, 2019). Findings from studies conducted with oral only tasks have also failed to find a clear positive influence of TR on NoM strategies (García Mayo & Imaz Agirre, 2016; Lázaro Ibarrola & Hidalgo, 2017), on the contrary, TR did lead to less L1 use (Azkarai & García Mayo, 2017). Therefore, it is anticipated that YLs in our study will not generate more focus on form on Day 2 than on Day 1, but they might resort less to PKL. Secondly, individual DG studies examining the impact of a pretask FFI with adults have demonstrated a positive influence on target form accuracy (Shintani et al., 2016), but not so much on general accuracy (Khezrlou, 2019). As far as YLs are concerned, findings from individual dictogloss have failed to show a positive impact (Gorman & Ellis, 2019). Conversely, pretask FFI has been occasionally used in adult collaborative DG, resulting in a positive effect on the number and focus of LREs (Kim & McDonough, 2011; Swain & Lapkin, 2001). Yet, low proficiency has also been claimed to hinder to those benefits (Leeser, 2004). As a result, we expect that YLs in our study will benefit from a dedicated FFI instruction especially conceived for children and that this pretask stage will positively influence the number of form-related LREs.

RQ1d: Less attention to the target forms (3S and POSS) than to other grammatical forms is expected, due to the that fact that the DG task is very much open to different interpretations from them (as explained before, this task is not placed at neither extreme of closed-open task design continuum, Eckerth, 2008). Yet, the design of the DG texts is foreseen to play a role, as more discussions are expected about the 3S when children are performing DG1 (seeded with instances of this feature) than when they are carrying out DG2. Likewise, more discussions are expected about POSS when children are performing DG2 than when they are carrying out DG1. Moreover, FFI is anticipated to encourage more attention to these forms than the canonical collaborative DG. Finally, the different nature of these features should not be overlooked. Owing to the greater salience and communicative value of possessive determiners, it is likely that learners will more easily concentrate on POSS rather than on the 3S (which is a non-salient and low communicative value feature) (see below for a further explanation of the target form features).

RQ2a: YLs' dyadic interaction has received increasingly more attention during the last years. In the case of oral tasks, Butler and Zheng (2015) showed that children aged 9-10 and 11-12 displayed mostly a collaborative pattern, but it was especially evident in the case of the older children. With regards to CW tasks, García Mayo and Imaz Agirre (2019) reported findings in line with Butler and Zheng's only in the case of researcher-selected dyads. Finally, Azkarai and Kopinska (2020) assessed 11-12-year-old children's patterns of interaction in a DG task and concluded that they were mostly cooperative (i.e. showing a high equality, but a low mutuality). Therefore, their results contradict the two previous studies. As the task used in this dissertation is a DG and our sample resembles that of Azkarai and Kopinska's, a cooperative pattern is mostly expected, probably limited by learners' scarce resources in the L2. Besides, TR has been shown to positively impact on YLs' patterns of interaction. García Mayo and Imaz Agirre (2019) examined the dyadic interactions of two age groups (7-8 and 8-9) while they carried out an oral task twice. At T1 older children were mostly classified as passive-parallel, but at T2 most of them were considered collaborative. In the case of the younger children, they experienced a slight change towards a more collaborative dynamic, although they were mostly assigned to this pattern at T1. Thus, TR is foreseen to help child learners to display a more collaborative interaction on Day 2.

RQ2b: CW has been consistently shown to lead to more benefits than solitary writing in a number of writing dimensions: (i) higher accuracy in teenager EFL learners (Basterrechea & García Mayo, 2013; Villarreal & Gil-Sarratea, 2019) and adult ESL (Storch, 2005) and SFL (Fernández Dobao, 2012; Gallego, 2019); (ii) higher fluency in adult SFL (Fernández Dobao, 2012; Gallego, 2019, but see Storch, 2005 for evidence of shorter collaborative texts); (iii) greater lexical variety and more grammatical complexity in adolescent EFL (Villarreal & Gil-Sarratea, 2019); and (iv), higher holistic measures as content and organization in adult (Shehadeh, 2011) and teenager EFL (Villarreal & Gil-Sarratea, 2019). One of the studies in the literature which failed to find any differences between collaborative and individual writing acknowledged that the benefits of the former over the latter may only arise provided that a collaborative pattern exists among the learners (Watanabe, 2019). Therefore, those learners working in pairs in the DG task are expected to produce a better writing than those who work individually. Regarding the impact of TR, although scholars examining individual writing have reported benefits (Amiryousefi, 2016; Khezrlou, 2019; Nitta & Baba, 2014), the few studies that have analyzed its impact on collaborative writing (Hidalgo & Lázaro Ibarrola, 2020; Porral Combarros, 2017) show mixed findings, as it has been claimed that evidence of L2 writing improvement is also dependent on the type of measurements used. However, as both studies employed unfocused tasks (more specifically, a picture narration task), we expect that the combination of TR with DG, a more focused task, could bring about more changes in children's writing, especially in the domain of grammatical accuracy. To conclude, written accuracy has also been claimed to be positively influenced by the provision of pretask FFI, as shown by a study with individual DG and adult EFL learners (Khezrlou, 2019). However, the impact of FFI has not yet been confirmed in the case of child ESL learners performing the same task (Gorman & Ellis, 2019), although this finding could have been affected by an unsuitable choice of the target form. Hence, given that our target forms (3S and POSS) have been carefully selected for the child EFL population with L1 Spanish, a positive impact of FFI on written accuracy is expected.

RQ2c: Finally, regarding the impact of collaboration on individual writing, Calzada and García Mayo (2021c) examined to what extent one collaborative dictogloss could influence individual YLs' writing in terms of CAF. However, they could not find any significant changes in any of the writing dimensions. Another study analyzing the impact of a longitudinal collaborative continuation task treatment (Qiu & Lee, 2020) did find changes in the holistic ratings of the individual writings completed before and after the instruction. Learners produced longer, more accurate and slightly more complex texts. Therefore, the repetition of the dictogloss task will probably allow learners to improve certain aspects of their individual writing. Although both learners performing the task individually and in collaboration may benefit from rewriting a text, the collaborative groups are expected to take a greater advantage of it thanks to their collaborative dialogue.

RQ3a: In line with a previous exploratory study (Calzada & García Mayo, 2020b), it is foreseen that YLs in this study will show awareness of the difference between L1 and L2 writing (reflected on their different degree of self-efficacy and anxiety). Furthermore, given children's natural instinct for social interaction and play (Halliwell, 1992; Pinter, 2015), we also foresee that they will display positive attitudes towards pair work already before the experimental procedure.

RQ3b: Three previous studies have investigated children's attitudes towards the DG task: Shak (2006), Calzada and García Mayo (2020b) and Kopinska and Azkarai (2020). Shak examined ESL children's perception by means of a questionnaire after carrying out an individual DG task twice, and reported that the second day learners had a more positive view of it (supporting the claims that task familiarity is associated with more motivation, Johnson, 2008). Calzada and García Mayo (2020b) assessed EFL learners' task affect (Lambert, 2017) after carrying out a DG task once, which some of them completed in pairs and others in small groups. Regardless of their group setting, all learners displayed a strong liking for the task, but were less aware of the learning opportunities it offered. Finally, Kopinska and Azkarai (2020) examined YLs' task motivation and views on individual and collaborative DG at the beginning and the end of a school year. They found that, throughout time, learners' task motivation towards collaborative DG was slightly higher than towards individual DG. Nonetheless, children's preference for each working mode remained relatively stable, as did their most and least preferred aspects of the DG task. Besides, posttask motivation always exceeded pretask motivation in both working modes.

Hence, in the present study child participants are expected to display a positive attitude towards the DG task, especially in the case of learners who experience it collaboratively. We foresee that learners will not object to repeating the task, as it has been show that their motivation levels are not negatively affected even after repeated encounters with the task. Finally, despite the lack of previous studies on the topic, providing learners with FFI prior to the task could possibly enhance their awareness of the linguistic benefits of the task.

RQ3c: YLs' linguistic repertoire does not consist of isolated languages which work independently, but it rather resembles more an intertwined network (Cook, 1992). What is more, previous research (Manchón et al., 2009; Pae, 2018) points to the fact that in order for a positive transfer between the L1 and L2 to occur, learners need to have attained a certain proficiency level in L2 (i.e. evidence of a linguistic threshold, Alderson & Urquhart, 1984; Clarke, 1980; Yamashita, 2001), although the impact of L1 writing skills on EFL writing has also been shown to be higher at lower proficiency levels (Schoonen et al., 2011). Therefore, in the case of low-proficiency child learners, L1 writing skills are not expected to exert a great influence on L2 writing skills. In addition to L1 writing, previous research has also demonstrated how learners' perception of writing influences their EFL writing quality (Pae, 2008). Thus, those child learners with a better attitude towards writing are anticipated to score higher in the L2 writing test.

5.3 Participants and setting

In order to answer our research questions, data were initially gathered from one hundred learners from two different state schools (School A and School B) in the same city, located in the BAC. In both schools, learners attended the 6th year of primary school, the last year of this education stage in Spain (ages 11-12). However, upon consultation with the learners' tutors while carrying out the experiments, the data from seven participants were discarded due to different reasons. Five learners had behavioral problems, another subject had a language specific impairment and, finally, another one was a native speaker of English who had just arrived in Spain. In summary, there were ninety-three (n = 93) children involved in the present study. The data reported henceforth correspond to this sample. The participants' names were substituted for an identifier in order to ensure their anonymity.

Before the experiment took place, all children except for two participants who were absent completed a language background questionnaire in their L1 (see A1 for the Spanish version and A2 for the English version in APPENDIX A), in order to gather information about their age, gender and language repertoire. They were aged 11-12 (M = 11.12, SD = 0.36). There were 48 male (52.7%) and 43 female (47.3%) subjects. As far as their language repertoire is concerned, the great majority of learners (93.4%) declared Spanish as one of their home languages[12], and all of them were able to read, speak and write in this language. Furthermore, when considering other social contexts in which Spanish was present, all stated that they used this language to communicate with their friends and 93.4% stated that they spoke Spanish at school. Finally, when

[12] The term "home language" is employed as many children declared that the language that they used with one parent could be different from the one they spoke with the other. Likewise, the language they used with their parents could be different from the one they used with their siblings. Hence, "home languages" is a more generic term than "mother tongue".

asked about the language they dreamt in (which reflects an inner use of the language, more related to the L1(s) than to the L2, according to Cook, 1998), 80.2% claimed that they did so in Spanish (16.5% could not provide an answer to this question).

Yet, despite the predominance of Spanish, it should be noted that learners lived in the BAC, where both Basque and Spanish are official languages, and moreover, most of them displayed a multilingual linguistic background. With regards to Basque, all children in this study attended the D model in their schools. This linguistic model involves learning in a Basque immersion setting, where all subjects are delivered in this language except for Spanish and English. Nonetheless, only 4.4% claimed to have only Basque as their home language, while the percentage raised to 29.7% in the case of children who had both Spanish and Basque as home languages. In this sense, the latest sociolinguistic report (Soziolinguistika Klusterra, 2017), which analyzed the presence of Basque in the city where the present study was carried out, showed that the use of this language in street communicative exchanges represented only 3.7% (vs 92.5% of interactions in Spanish), although it raised to 7.5% taking into account interactions among children in the region of Álava (2-14 year-olds), where the city of this study is circumscribed. As a result, participants can be quite confidently described as L1 Spanish despite attending a D linguistic model which immersed them in Basque.

As to foreign languages, starting with English, 12.1% acknowledged they used this language at home. When self-assessing their competence in English, 89% claimed to be able to read and speak, and 83.5% declared so in the case of writing. It is interesting to note that, as we will see below, this percentage was not higher taking into account that they had had at least 6 years of EFL instruction.

Finally, regarding foreign languages other than English, 19.8% claimed to use them as home languages, and an even higher percentage (27.5%) claimed to be able to speak them. These rates decreased when children had to self-assess whether they were able to read or write in any of these languages (16.5% and 13.2%, respectively), indicating that, most probably, they learnt them at home from their parents, but did not receive any sort of formal instruction.

In both schools, learners at the time of the study received three hours of EFL, where they followed a communicative syllabus based on the manual by Read and

Ormerod (2018). In this coursebook most attention was devoted to learning vocabulary on daily topics and learning basic grammatical structures with a functional or pragmatic perspective. Furthermore, in School A, the English teacher had implemented a project-based approach in addition to the content in the book, where children had to work individually or in groups on a given topic and present it orally in class with the help of slides. In addition to the information that we gathered about the EFL subject from each of the schools, we asked children about their EFL trajectory in the language background questionnaire. The average years of English study was 7.26 (SD = 1.53), which approximately equaled their number of years of formal education in English in the national curriculum since the last year of pre-primary school (ages 5-6). Moreover, the average hours of English study per week was 3.77 (0.93). In fact, a considerable number of children claimed to receive extracurricular classes (43.3%). To conclude the language questionnaire, children were asked whether they studied other foreign languages at the moment, and the great majority (75.8%) denied doing so.

Our research design involved splitting the participants into three different groups: a collaborative group (Collab), who performed the DG tasks collaboratively in pairs, a form-focused instruction + collaborative DG group (FFI+Collab), who received a pretask instruction on the target forms (3S and POSS) prior to completing the DG tasks in pairs, and finally, a comparison group (Comp), who performed the DG task individually. All three groups engaged in pretesting and posttesting (more information about tasks and procedures is provided below). Assigning participants to one of the three groups, unfortunately, could not be done completely randomly due to different reasons. Firstly, given practicality issues involved in delivering a pretask form-focused instruction in a limited a number of sessions, all the subjects from the FFI + Collab group came from School B. Furthermore, most subjects from the Comp group were learners whose parents had provided permission for participating in the study, but not for recording their interaction. Consequently, and following Loewen and Plonsky's terminology (2015), our research should be identified as a quasi-experimental study. The final distribution of participants across the three conditions was the following one: Comp (n = 34), Collab (n = 28) and FFI+Collab (n = 31).

Before the experiment, all learners except for two that were absent took a *Flyers* standardized proficiency test (Cambridge Assessment English, 2018). The test aimed to provide evidence of a complete basic command of English, corresponding to an A2 level in terms of CEFR. They completed the Reading and Writing paper (without Part 7, writing a story) and the Listening paper. Considering that learners were going to be tested primarily on writing and that we were interested in their metalinguistic reflections, we convened to assign the former a weight of 70% of and the latter 30% over the total proficiency score. After checking for normality in the distribution of the data (Q-Q plot of standardized residuals) and homogeneity of variance between the groups (Levene's test, $p = 0.091$), we ran a one-way ANOVA in JASP (JASP Team, 2019) in order to ascertain the equivalence between the three groups before the experimental procedure took place. As suggested by Larson-Hall and Plonsky (2015), we focused on the means differences confidence intervals (CI) resulted from the ANOVA multiple comparison post-hoc Tukey test. Table 3 summarizes the descriptive statistics and Table 4 displays the inferential results:

Table 3 Descriptive results for the proficiency test in each condition

	M (SD)	**25th Q**	**50th Q**	**75th Q**
Comp (n = 33)	39.73 (26.84)	14.74	33.59	56.99
Collab (n = 28)	50.29 (19.86)	35.37	51.88	66.69
FFI + Collab (n = 30)	36.99 (21.72)	20.89	33.12	42.65

Table 4 Confidence interval mean difference between the three conditions for the proficiency test

	95% CI	***p***	**Cohen's *d***
Comp vs Collab	[-24.77 - 3.66]	$p = .19$	-0.44
Comp vs FFI + Collab	[-11.21 - 16.70]	$p = .89$	0.11
Collab vs FFI + Collab	[-1.23 - 27.84]	$p = .08$	0.64

On average, learners in the Collab group scored highest in the *Flyers* test, followed by the Comp and FFI+Collab. Yet, regarding inferential statistics, we failed to find any differences between the groups, as the CI for mean differences cross the value of 0 in all three cases. The greatest difference could be found in the comparison between Collab vs FFI+Collab, as, with 95% confidence, the lower bound value (-1.23) strays across zero for a very short distance. However, the interval is just as wide as in the other two comparisons, showing a potential difference in the proficiency test of as large as 27.84. All *p*-values reflect a lack of statistically significant difference ($p > .05$). Finally, in the same line, the effect sizes, according to Plonsky and Oswald's (2014) benchmark for inter-group mean difference contrasts, range from very small ($d = 0.11$) to small-medium ($d = -0.44$ and $d = 0.64$).

These results were translated into *Flyers* official scores, which are interpreted on the basis of shields representing the degree of completion of CEFR benchmarks (Cambridge Assessment English, n.d.-a). In general, learners did not show a complete elementary level (A2), but they were rather closer to an A1 level. This was especially evident in the FFI+Collab group, as 75% of the learners scored less than 42.65 in the *Flyers*, that is, below our own "pass" threshold. In contrast, the Collab group showed the strongest performance in the test, as the last two quartiles obtained a pass mark or higher.

5.4 Materials and procedure

5.4.1 Instruments

5.4.1.1 The dictogloss task and the target features

In the current dissertation the DG task was chosen as the literature has demonstrated that it had been the most widely used task in CW research (Storch, 2016, 2019). Defined as a focused task, that is, one in which the researcher can set some linguistic goals that are usually not unveiled to the learners, DG is considered one of the most effective CW tasks at fostering *languaging* and FonF opportunities. In fact, recently, DG has also recently been proved effective for drawing YLs' attention to form (Calzada & García Mayo, 2020a, 2021a), although some limitations have been reported regarding its potential for generating discussions about target features.

Moreover, in these studies some task features were indicated to be key for DG to succeed: careful choice of the text vocabulary, task familiarity and conscious selection of the target forms taking the learner population characteristics into account. The DG texts used in this study had been previously employed in other experiments with YLs in

EFL settings, but never in combination as part of the same pedagogical intervention. The two texts (DG1 and DG2) target two different features (3S and POSS, respectively), which have been found to be problematic for EFL learners from a linguistic and pedagogical perspective (Spada & Tomita, 2010), and especially, for Spanish L1 learners. The main characteristics of these linguistic forms will now be explained, together with a review of the different pedagogical proposals that were explored in research seeking to facilitate their acquisition.

5.4.1.1.1 The third person singular present tense marker -*s* (3S)

In SLA studies, the 3rd person singular present tense marker -*s* (3S) has received a great deal of attention, starting from the morpheme studies up to nowadays. The morpheme studies are characterized by learner spontaneous data gathered cross-sectionally, which is usually analyzed in terms of accuracy of suppliance. Dulay and Burt (1973, 1974) conducted their first empirical study on the L2 acquisition order of the grammatical features of English by L1 Spanish and Chinese learners, and claimed that L2 acquisition follows a quite systematic path regardless of the learners' L1. According to their sequence, 3rd person -*s* would be placed towards the end of the acquisition process (specifically on the seventh position), preceded by (i) plural -*s*, (ii) progressive -*ing*, (iii) copula *be*, (iv) auxiliary *be*, (v) articles and (vi) irregular past tense, and followed only by the Saxon genitive *'s*. In fact, despite the order mentioned above being slightly different from that found for the same morphemes in L1 acquisition research (Brown, 1973), their results provided evidence in favor of the existence of universal cognitive mechanisms which enabled learners to discover the structure of a particular language (Hironymous, 1993, but see Luk & Shirai, 2009 for counterevidence).

Yet, a study by Stauble (1984) did reveal an influence of the learners' L1. The author examined six Spanish and Japanese learners' suppliance of the 3S, along with other morphemes, in conversation data. Although the 3S accuracy was much lower than for other morphemes (e.g. past irregular), the results revealed that the two Spanish learners (described as mesolang speakers, that is, at a medium stage of interlanguage development) achieved higher levels of accuracy than the two Japanese mesolang learners (56% vs 10-19%).

Notwithstanding the evidence showing a late acquisition of this morpheme, moderated to a certain extent by the learners' L1, SLA research has not reached a consensus about its complexity. This divergence of opinions originates, as explained by Spada and Tomita (2010), from two ways of describing complexity. On the one hand, pedagogical descriptions of grammatical rules describe "complex" or "simple" features in terms of how difficult they are to explain (i.e. their teachability). For instance, according to Krashen (1982), 3S is a simple form because of its paradigmatic uniqueness. On the other hand, psycholinguistic accounts like the one defended by Goldshneider and DeKeyser (2005), drawing on N. Ellis' (2006) associative learning theory, take into consideration other morpheme factors. For example, N. Ellis (2006) identifies up to three factors contributing to making morphemes like 3S complex forms:

- multifunctionality: it can indicate plurality, possession, as well as occurring in contracted form in copula and auxiliary *be.*
- redundancy: the person is indicated by the obligatory pronoun or noun in English, hence adding low functional value.
- aural perception difficulties: (i) the low number of phones (i.e. phonetic substance) and the variant allomorphs [s] and [z] and [əz]; (ii) the morpheme does not constitute a syllable as it lacks vowels; and (iii) it is a sonorous morpheme, but it frequently occurs in complex codas in final position of verbs in combination with other consonant sounds (e.g., /-lz/ in /ˈtrævəlz/ (travels), /-ks/ in /laɪks/ (likes), or /-vz/ in /bɪˈli:vz/ (believes)).

Furthermore, N. Ellis (2006) introduces the cognitive concept of "blocking", which can help understand the complexity of 3S. Blocking occurs as a result of "automatically learned inattention" (p. 178), which can, indeed, be long-lasting and attenuate any further learning about that form. When two linguistic cues realize the same meaning (in this case, the 3S and the subject of the verb), the more salient one (the subject) overshadows the other. Moreover, when inattention to the *-s* becomes systematic, it is difficult to reverse it.

From ISLA research, there have been multiple pedagogical proposals to direct learners' attention to this particular feature. In fact, it has been demonstrated, at least in the case of adult EFL learners, that incidental learning of 3S from exposure (even when this is maximized) is not likely to take place (Loewen et al., 2009) and, therefore, a dedicated FonF instruction is considered necessary. In this sense, some studies have investigated how, within a communicative context, using tasks that foster noticing (Schmidt, 1990, 2010) and explicit knowledge of the form can help to improve learners' implicit knowledge (i.e. known as the *weak interface* in cognitive psychology, see R. Ellis, 2018). Empirical research addressing this topic will now be reviewed, first, from TBLT and, secondly, from CW with regards to adults and YLs.

Mostafa and Kim (2020), for example, compared how input- or output-based tasks could help to automatize explicit knowledge of 3S*.* In Suzuki and Dekeyser's view (2017), automatized explicit knowledge is accessed as rapidly as implicit knowledge, but these two types of knowledge differ in the fact that the former involves awareness of the target forms whereas the latter does not. Seventy-two adult ESL learners (age M = 24.7, SD = 6.88, different L1 backgrounds) took part in Mostafa and Kim's study and there were two different treatment groups. One of them followed Processing Instruction, that is, input-based activities (such as picture and sentence matching, where the sentences contained instances of the target form). The other group was administered output-based instruction (using the same pictures, the learners had to write down sentences which would trigger the use of the target form). There was also a control group that did not receive any sort of instructional treatment. Apart from the different instruction, both the input- and output-based groups received a short explanation about the target form rules. Using a range of tests to measure explicit (error correction) and automatized explicit knowledge (oral narrative), the findings revealed that the two treatment groups were able to improve significantly the two types of knowledge of the 3S from the pretest to the posttest and to the delayed posttest, while the control group failed to show any significant improvement. Regarding intergroup comparisons, the output-based group significantly outperformed the input-based group only in the oral narrative task posttest, that is, in the development of automatized explicit knowledge. Yet, this difference disappeared in the delayed posttest.

Shifting now to YLs, Roos (2019) explored the use of form-focused communicative tasks with German L1 learners of English from two school years (grade 6, n= 6, ages 11-12, and grade 7, n = 6, ages 12-13). Furthermore, taking a Processability Theory approach (Pienemann, 1998, 2005), she investigated the impact of that instruction depending on the learners' readiness to be taught 3S. The treatment took place over the course of two weeks in four 20-minute lessons. During those lessons, the teacher never made the children aware of the focus of the tasks, but did bring up the rule while they were completing them (a popular mnemonic rule used in German primary EFL). An example of a task used during the treatment was a typical oral information-gap task, where learners had to inquire of each other about the daily schedule of a particular character. Using informal interviews as pretest, posttest and delayed posttests the author calculated the use of the *-s* in obligatory contexts. The results of the pretest were used to classify learners as developmentally ready/not ready to acquire the form. The findings showed that the use of 3S increased from the pretest to the posttest, especially in the case of four learners who had demonstrated to be "ready" to acquire it in the pretest. The results of the delayed posttest also showed an increase for half of the learners, while two children performed slightly worse than in the posttest, but still better than prior to the FFI. Hence, the author suggests that the instruction period helped all learners regardless of their developmental stage, supporting learners "in gaining control of a feature that had already emerged in their interlanguage" (p. 294).

Turning now our attention to studies using CW to focus learners' attention on 3S, Basterrechea and García Mayo (2013) compared CLIL and EFL adolescent learners' (n = 81, ages 15-16, B1 intermediate proficiency) performance in a collaborative DG designed to target this feature. In contrast to the aforementioned studies, instead of using a pretest-posttest design, the authors were interested in the learners' metalinguistic deliberations during DG, as well as its effect on 3S written accuracy. The analyses of the LREs showed that almost half of the discussions held by CLIL learners were devoted to this form, whereas in the case of mainstream EFL learners they represented slightly less (41.7% of all LREs). What is more, in both learning contexts, correctly resolved LREs about 3S were more frequent than incorrect or unresolved deliberations. Finally, with regards to accuracy, CLIL dyads outperformed EFL learners. These results were compared to those of a control group that performed DG individually. In the case of EFL, there was no significant different in accuracy, but the opposite was true in the CLIL setting.

As far YLs are concerned, Calzada and García Mayo (2021a) took the same approach as the aforementioned study, that is, to analyze LREs that originated during dictogloss. Instead of examining the accuracy of the collaborative texts, the authors tallied the resolved instances which were incorporated into the joint text. Contrary to Basterrechea and García Mayo (2013), the findings revealed that YLs did not focus on the target form to the same extent as on other grammar forms. In fact, LREs about 3S only represented 6% of all languaging episodes. Moreover, episodes on mechanics (spelling and punctuation) were slightly more frequent than grammatical LREs (40% vs 37%). However, despite the low number of target form LREs, children were able to correctly resolve most of those deliberations, and their oral resolutions were then largely incorporated into the joint text (81% of the times).

After reviewing the most relevant empirical pieces of research regarding task-based interventions targeting 3S, some general conclusions can be drawn. Tasks, especially output-based collaborative tasks, offer opportunities for noticing this linguistic feature to both adult and child learners. In fact, YLs who have not acquired the form can benefit from this type of instruction, but also those who already have mastered the form, as they can consolidate their knowledge, for instance, by providing feedback and clarifications to their peers. When it comes to CW, adults seem to benefit more than child learners from collaborative DG targeting 3S. Although the two reviewed studies are not directly comparable (different proficiency levels, different texts used, etc.), we could wonder if YLs need extra support in order to attend to this feature during DG. In fact, their inclination towards meaning and communication rather than form could be hindering some grammar learning opportunities. Thus, in the present dissertation we aim to explore the impact of a pretask FFI on children's attention to 3S and subsequent written use of this form.

5.4.1.1.2 Third person singular possessive determiners his/her (POSS)

If developmental and psycholinguistic aspects were most central for explaining what hinders the acquisition of 3S, in the case of the possessive determiners his/her (POSS) the difficulties may be explained by the learner's L1 influence. In fact, the acquisition of these forms has been reported to be especially complex for Spanish L1 learners due to several L1-L2 differences. These differences are summarized by Sato and

Loewen (2018): in Spanish, in contrast to English, there is only one third person possessive determine *(su)*, which agrees in number with the possessed entity *"su perro"* (his/her or their dog), *"sus perros"* (his/her or their dogs), and not with the gender of the possessor. In a nutshell, there is an additional processing load for these speakers as they need to assign a biological gender to the possessor pronoun which may or may not agree with the L1 grammatical gender of the possessed entity. In order to illustrate this processing complexity, the authors provide the following example (p. 12):

(7) *Ella le pidió a su padre que le cambiara su pelo*
(She$_i$ asked her$_i$ father to change her$_i$ hair)

In this example, the subject is female, but the possessed entities have, respectively, a male biological and grammatical gender (provided that there is no prior reference to "he" or any male reference). For an L1 Spanish speaker, due to the L1-L2 differences explained above, it would be common to say in English: *She$_i$ asked his$_i$ father to change his$_i$ hair. In fact, research has shown that Romance language speakers (Spanish, French, Catalan) follow a developmental pattern in the acquisition of these forms (J. White et al., 2007). During the first stages (pre-emergence stage), instead of assigning the L1 gender of the possessed entity to the determiner, learners avoid using the form *(*She asked father to change hair)* or tend to use the definite article *(*She asked the father to change the hair).* In later stages (emergence stage) , learners show preference for one of the possessive determiners. Previous research conducted by Muñoz (1994) with Spanish/Catalan bilinguals demonstrated that there was an increase in the number of errors in POSS as the learners' proficiency increased.

In addition, the possessive determiner distinction poses problems for learners of L1 backgrounds different to Romance languages. Since the sample in the present dissertation are also Basque learners, the differences between this typologically different language and English have also been explored in the work by Imaz Agirre and García Mayo (2013). In Basque, there are inflected possessive determiner phrases (DPs), in which genitive suffixes (expressing the relationship of possession) may co-occur with the definite article (p. 418):

(8) Ber-e liburu-a.

DEM-GEN BOOK-DET

'His/her' book

Unlike English and Spanish, there is no expression of grammatical gender in nouns or determiners (except for special cases of loan words). However, possessive determiners do agree in number (singular/plural) with the possessor. As in the following example (p. 418):

(9) a. Jon ber-e ama-rekin jolaste-n ari da.

John DEM-GEN mother.DET-with play-PRES PROG AUX

'John is playing with his mother'

b. Jon eta Mikel ber-aien ama-rekin jolasten ari dira.

John and Michael DEM.GEM-PL mother.DET-with play.PRES PROG AUX

'John and Michael are playing with their mother.'

Although one could hypothesize that this possessor-determiner agreement could be of help for L1/L2 Basque learners of English, Imaz Agirre and García Mayo (2018), in their comparison of advanced and intermediate adult learners of English (considered as balanced Spanish/Basque bilinguals), demonstrated that the same developmental pattern shown for Romance speakers by White et al. (2007) applied. In other words, balanced bilinguals seemed to rely on the Spanish gender of the possessed entities when they produced the possessive determiners in English. In fact, this occurred both in gender matched (when the possessor's biological gender in English and the possessed entity's grammatical gender of the equivalent Spanish word coincided), but also in gender-mismatched contexts (although to a lesser extent). Finally, the authors reported that the most problematic contexts were those in which the possessor and the possessed entity were animate nouns.

Despite not being as widely studied as 3S, from the field of ISLA some pedagogical interventions have been put to practice to explore how the acquisition of his/her can be facilitated. In a meaningful context, possessive determiners, in contrast to the third person singular present tense marker, have a higher communicative value. As explained by Sato and Loewen (2018), while omitting 3S during interaction may not provoke a communication breakdown, errors related to POSS (omissions or agreement errors), may. Indeed, the wrong use of these forms can lead to face threatening situations, and hence, learners' stronger motivation to gain a better control of them (J. White, 2008).

Some of the pedagogical interventions carried out with YLs will now be reviewed, starting from task-based studies and moving on to CW. J. White (2008) reviews three studies targeting his/her in three different ways, but we will only concentrate on Study 1, as it is the only one where form-focused instruction was used (the rest of the studies were closer to pure structural activities and grammar translation). In this study, set in the ESL context in French-speaking Canada, the researcher assigned learners (age 12) to two treatment groups: the input enhancement group, known as E (n = 30) and input enhancement + extensive reading and listening group, known as E+ (n = 27). Finally, there was a control group that read all the enhanced versions of the texts in common in E and E+. To guarantee the meaningful nature of the instruction, learners were never presented the rule of the target form, but noticing was also fostered by the reading comprehension questions, which required a correct understanding of the possessing relationships.

The treatment spread out over the course of two weeks, with 30-minute sessions each day, and before and after the treatment the learners were administered a pretest, posttest and delayed posttest consisting of a measure of explicit knowledge (error correction) and implicit knowledge (oral narration). The findings from the posttest revealed that the treatment effect was strongest for E and E+, but in the delayed posttest these two groups did not perform differently from the control group. Moreover, the initial boost in the posttest was especially strong for E+. These results led the authors to claim that this input-based implicit type of instruction was inadequate for the target population, as it did not make use of children's analytic skills and problem-solving abilities (this is the main reason why Study 2 and Study 3 opted for more explicit techniques).

A study which did show to a larger extent YLs' problem solving skills was that by Calzada and García Mayo (2021b). As in a previously reviewed investigation where these authors targeted 3S by means of a collaborative DG task, in this study they administered the same type of task to sixty-six EFL children (ages 11-12, A2 elementary proficiency) to facilitate their acquisition of POSS. Before the experimental dictogloss, the learners were familiarized with the task through an individual DG. Furthermore, the text to be reconstructed was carefully designed so that all instances of his/her (i) had animate possessors and possessed entities (Imaz Agirre & García Mayo, 2013), (ii) they had a clear communicative value in relation to the gist of the story. All pairs were videorecorded during the task performance and all LREs were tallied and classified in terms of topic, resolution, incorporation to the joint text and depth of engagement. The results showed that on average learners focused significantly more on other grammatical forms than on POSS. Furthermore, simple engagement predominated in the LREs related to grammar, including POSS. Yet, as in Calzada and García Mayo (2020b) children were able to correctly resolve most of the deliberations, and incorporated these correct decisions to a larger proportion than incorrect LREs in their joint texts.

In summary, research on pedagogical interventions targeting POSS in early L2 English contexts shows that, although mere input is not sufficient, engaging YLs with communicative output-based tasks can be more helpful in creating noticing opportunities. However, the low amount of target form-related LREs shows that it might be necessary to reinforce collaborative DG with other form focused strategies. This is, indeed, what we aim to explore with the inclusion of pretask FFI stage.

5.4.1.1.3 Dictogloss text characteristics

Two dictogloss texts were used for the present study. One of them, “Naughty Laura” (referred to as DG1) was seeded with 3S instances, and the other, “A celebration” (labelled DG2), targeted POSS (see the texts A3 in APPENDIX A). Both dictogloss texts had been previously used in our research with YLs (see Calzada & García Mayo, 2020b and Calzada & García Mayo, in preparation). From these preceding studies and the piloting phase (explained below in more detail), we concluded that they were suitable for children with an A1-A2 proficiency. Both texts took the form of a narrative story, the genre which learners are most familiar in primary education in Spain (Fernández et al., 2019).

In addition, in consultation with the children’s English instructor, we made sure that participants in the present study were also familiar with the vocabulary and the topics of the texts, namely, family, school and food vocabulary. Furthermore, the texts were recorded by a highly proficient L2 English speaker so as to avoid any change in the reading pace throughout the different interventions as a result of different researchers reading the stories aloud. In order to avoid any performance differences which were due to text characteristics and not to the target form itself, a range of text-based measures[13] was calculated (summarized in Table 5), which proved their similarity:

Table 5 DG1 and DG2 text characteristics

	Flesch–Kincaid	GI	Words	Recording time	Words/ min	Target form instances	Clauses/ T-unit	Coord/ T-unit	MLC
DG1	4.9	6.52	122	01:17	95.1	15	1.05	0.55	6.42
DG2	3.5	7.09	127	01:18	97.7	8	1.12	0.44	7.05

Regarding the selection of target feature forms, in the case of DG1, we made sure that all the verbs had a regular present tense third person singular agreement (i.e. irregular spelling forms such as ‘does’ or ‘watches’ were avoided). As for DG2, these forms were controlled so that they always referred to animate possessing and possessed entities. Thus, there were two instances of masculine possessor and masculine possessee (his), two instances of masculine possessor and feminine possessee (his), two instances of feminine possessor and feminine possessee (her), and two instances of feminine possessor and masculine possessee (her). In sum, there were eight instances of POSS. The lower number of target form instances in this text as compared to DG1 is due to the difficulty of seeding possessive determiners in a text of the equal length while keeping the coherence of it. Moreover, as reviewed previously, 3S, apart from being a bound morpheme, is considered a phonologically low salient form and has a lower communicative value than possessive determiners. Therefore, we assumed that having more instances of the present tense marker would be necessary for children to be able to notice it.

[13] The choice of clauses/T-unit, coordinated clauses/T-unit and MLC is explained in 5.5.4.1 Text-based measures.

As far as lexical (GI) and grammatical complexity (Clauses/T-unit, Coord/T-unit, MLC) measures are concerned, both texts display comparable values. The texts were designed taking into account child learners' potential command of compound structures (coordination and subordination). As subordination is not usually present at this stage (Mylläri, 2020; Torras et al., 2006), it was decided to reduce as much as possible the number of this type of clauses. As a result, the subordination ratio (Clauses/T-unit) is very close to 1 in both texts, implying that dominating structures in the text were independent clauses (i.e. those that are counted as separate T-units). In contrast, the coordination ratio (Coord/T-units) shows that almost half of the independent clauses were coordinated in the case of DG2, and a slightly more than a half in the case of DG1. Finally, the MLC ratio indicates an average of 6-7 words per clause, which contributes to an appropriate composition of the text given the fact that it was presented aurally.

Finally, in order to familiarize all learners with the DG task procedure, a similar task in Basque was designed (see A4 in APPENDIX A). The trial DG task targeted the Basque verb suffixes *-t(z)en* and *-t(z)eko*. Children carried out this task with their tutors in class some weeks prior to the experiment.

5.4.1.2 The pretask FFI

As part of the treatment of one of the experimental groups (FFI+Collab), two pretask focus on form instruction (FFI) mini-lessons were designed in order to raise YLs' awareness of 3S and POSS (see A5 in APPENDIX A). Two main reasons encouraged the use of a pretask preparation stage. Firstly, the thorough review of the previous studies using collaborative DG demonstrated that 42.8% of the studies employed a preparatory stage, of which 17.1% constituted FFI. Secondly, the results of previous studies exploring the impact of this task on YLs' performance (Calzada & García Mayo, 2020a, 2021a), had clearly shown that DG was not enough to direct children's attention to the target forms, especially in the case of those forms which were not salient and had a low communicative value (such as 3S) or for which children had received scarce previous instruction and feedback in their classes (such as the articles in English). The present pretask lessons were designed in order to suit this population and their appropriateness was checked in consultation with the learners' English teacher. For both forms, the mini-lessons consisting of three activities that were 10 minutes long were delivered by the author of this dissertation. Moreover, the sequence of activities was the same for 3S and POSS except for the last part (the production part). The activities were the following:

- 1st Part - Input enhancement: in this activity, learners were presented with a short narrative text in writing, similar in structure and containing the same number of target form instances as the dictogloss text they were going to listen to afterwards. Those instances were enhanced typographically to call their attention and facilitate noticing. In the case of 3S, the morpheme was substituted with a drawing of a snake which resembled the letter 's'. The size of the drawing was larger than the text font. On the other hand, in the case of POSS, all the 'his' instances were in a bold blue font, and the 'her' instances were in bold red. In addition, the drawings of the two characters of the story, John and Jenny, always preceded the possessive determiners. In this way, we aimed to facilitate the connection between the forms and their animate reference. Finally, after each possessive determiner, the possessed entity was underlined.
  Learners were required to read the text individually on their own. Afterwards, the researcher asked them a few questions about the content and vocabulary. Then, he read the text aloud emphasizing the target forms. Once finished, he asked them to reflect why some elements (the target forms) were larger and in bold. In summary, this activity can be considered as "non-interactive focus on form", in terms of Ellis' (2016) classification of FonF techniques. In his words, this type of strategy occurs when "when learners are asked to process oral or written input where specific target features have been highlighted" (p. 6).
- 2nd Part - Elicitation and provision of the rule: the researcher asked the learners to turn the page. The following activity required them to complete the "golden rule" which explained how the target forms in question worked. The researcher gave them 2-3 minutes to discuss it with their partners and afterwards he elicited some answers from them. Finally, he provided the rule and made sure everybody had understood correctly. In contrast to Roos (2019), we were not aware of any popular mnemonic

technique to retain 3S in the early Spanish EFL context, so the golden rule of the 3S was designed ad-hoc for this study. Yet, the golden rule of the possessive determiners was taken from IRIS repository materials (Marsden et al., 2016), more specifically, from the study by Horst et al. (2010). These authors had used this rule with French-speaking children aged 11-12 in the Canadian immersion context.

- 3rd Part - Production: in the last part of the worksheet, learners had to engage with a collaborative production-based activity. In the case of the present tense marker, learners had to add snakes to a short a text (one of DG texts used in Calzada & García Mayo, 2021a) when they considered it necessary. Conversely, in the last activity of the possessive determiner worksheet, learners had to fill in the gaps of a short narrative text adapted from the materials in Horst et al. (2010). After the children collaboratively discussed these activities, the researcher went through the text again providing the solutions and solving any doubts learners could express.

Finally, it should be noted that all the instructions in the worksheet were in English and Spanish. In this sense, although the researcher led the mini-lesson primarily in English, he occasionally shifted to Spanish for class management (keeping discipline) and to provide certain clarifications of the activities when learners had doubts.

5.4.1.3 The written languaging worksheet

In the case of the Comp group, after the children had performed the individual dictogloss, they were asked to reflect on the difficulties and doubts during the task and to write down on a separate worksheet, known as the individual languaging worksheet (see A6 in APPENDIX A). In order to help them enunciate those reflections, learners were provided with a layout which illustrated how their thoughts could be expressed in their L1. This model was inspired by the guided noticing materials developed by García Mayo and Loidi (2017) in their study with secondary school EFL learners (ages 13-16) from the BAC. However, our participants were also encouraged to use their own words if they felt more comfortable doing so. All learners regardless of their treatment group had previously received a similar worksheet some weeks before the experiment in the Basque preparatory DG task (explained above). The teachers were advised on how to encourage the metalinguistic reflection of their learners during that stage, both providing and eliciting examples of written languaging.

5.4.1.4 The writing tests

Three individual narrative L2 writing tests (see A7 in APPENDIX A) were used as a means to assess the learners' L2 writing skills prior to the treatment, and as a posttest and delayed posttest to compare the quality of their L2 written product after the two experimental dictoglosses. The writing activities can be classified as "continuation tasks", that is, learners were presented with the beginning of a story and a picture related to it. The researcher read that introduction aloud and resolved any understanding difficulties. This type of tasks were chosen as they were considered to offer more guidance and support than picture description or free writing activities, and they had been claimed to encourage more accurate writing performance (Peng et al., 2020). Moreover, we believed that the narrative genre, already extensively worked in the L1 and to a certain extent in the L2 by this stage in the BAC (Basque Government, 2015), could help child learners experience less anxiety during the tests. Finally, it should also be noted that this genre was aligned with the nature of the DG texts presented during the experimental procedure and, therefore, we intended to make these learners more aware of the narrative conventions.

The topics of the three writing tests were intended to be familiar to the participants. In "Writing in English 1" children had to continue a story related to an unexpected holiday situation; in "Writing in English 2" they had to write about a boy who lost the present for his best friend's birthday; and finally, in "Writing in English 3" they were required to write about a family day in the forest. All three stories dealt with what happened to a character and were framed in the present to encourage the use of 3S. Besides, they also introduced more female and male characters so that learners could resort to POSS at some point in the story.

The suitability of the stories was checked in a pilot study conducted in another school that did not participate in the final experiment. The learners in the pilot study were in the 5th year of Primary (aged 10-11, n = 20) and had a similar linguistic profile to those of the present study. However, as the learners were enrolled in a CLIL setting, they received a greater amount of English exposure hours at school (7 hours) than the learners in the experimental study. Therefore, we considered that notwithstanding their age difference, the pilot sample was comparable to the experimental one.

The researcher administered the different writing tests randomly to the learners before and after the pilot DG procedure. During the completion of the tests, all the questions raised by the children were written down, as well as anything that called the attention of the researcher, so as to improve the final version of these activities. Consequently, the writing space provided was considered excessive, as no child was able to fill in all lines at their disposal. Hence, this space was shortened in the final version. Moreover, some children expressed difficulties in understanding the topic of the story just by reading the beginning of it. For this reason, we included a general instruction at the top of the worksheet which summarized the gist of the text they had to write. Although children in the pilot study did not express that one of the tests was more difficult than another, the order of the tests was counterbalanced in the experimental study.

For the L1 writing test, three different test models were designed (see A8 in APPENDIX A). Following the same procedure as in the L2 writing tests, learners were required to continue a story after reading the beginning of it. To keep the resemblance with the L2 writing tests, the three stories were set in the present and had a third person character. Previous L1 writing development literature had shown that children by the age of 8-9 are already capable of significantly improving the cohesion and complexity of their stories, and by 10-11 years their narrative compositions can already be considered mature (Artiles & Jiménez, 2007; Yan et al., 2012). Furthermore, narration is the most frequent genre in school reading and writing tasks in Spain (Fernández et al., 2019). After piloting the three test models, we considered it necessary to extend the writing space in the worksheet, since most children were very fluent in L1 writing.

5.4.1.5 The attitude questionnaires

In order to tap into the learners' affect, we designed a pretask and a posttask questionnaire (referred to as AQ and DQ), which were administered before and after the DG procedure, in Week 1 and Week 4. Although in the L2 writing literature questionnaires have been more frequently used with adult learners (Aula Blasco, 2016; Fernández Dobao & Blum, 2013), there are recent examples where such instruments have been successfully employed with child learners (Calzada & García Mayo, 2020b; Kopinska & Azkarai, 2020). As suggested by Pinter (2011), surveys need to be carefully designed when we aim to investigate child learners' perceptions, and ideally, they should be triangulated with other sources, such as interviews, which were conducted in the form of focus groups at the end of the experimental procedure, in Week 6. In fact, incongruences have been reported between what child participants state in questionnaires and in interviews (Lamb, 2003). In the present dissertation, participants were older children who already had some experience in completing surveys, and we considered that adapting adult-like format questionnaires was sufficient. In what follows, the piloting of the two questionnaires is briefly explained.

Regarding AQ, it consisted of 44 Likert-scale items (19 of which were reverse coded) and 4 open-ended questions regarding five different constructs: (i) anxiety and self-efficacy in L1 writing (Spanish), (ii) anxiety and self-efficacy L2 writing (English), (iii) collaborative work vs individual work, (iv) collaborative writing vs individual writing, and (v) L2 learning. As in the case of the writing tests, the questionnaire was piloted in the aforementioned 5th year primary class. It mainly served as a way to know which items could be most confusing for the children and other aspects related to the layout of the questionnaire. Two versions of the Likert scale were used in the pilot stage: one including a numeric scale (1-5) and another one featuring different smiley icons. Although none of the versions was clearer than the other, smiley icons were selected for the experiment as this was deemed more attractive and motivating for children. The smiley scale was the same as in Tellier (2015), retrieved from the IRIS repository (Marsden et al., 2016), and the meaning of each face was directly translated into Spanish. Furthermore, although at the pilot stage the meaning of those smileys only appeared on the first page, it was considered it necessary to repeat that legend at the top of every page so that children did not have to go back and forth in order to recall the face meaning.

As far as the questionnaire length is concerned, due to the high number of dimensions we aimed to tap, AQ contained more questions than previous studies looking into child motivation and attitudes (Kopinska & Azkarai, 2020; Matsuzaki-Carreira, 2006; Pladevall-Ballester, 2019). However, other examples in the literature, (Anam & Stracke, 2016; Gallardo del Puerto & Blanco Suárez, 2021; Tragant et al., 2013) show that EFL participants as young as 9-13 are able to deal with relatively long questionnaires (containing more than 30 items). Despite our initial concerns about child participants' concentration capacity, during the piloting sessions we did not observe any sign of survey fatigue, and they were able to complete it well within 30 minutes. The definite version of the AQ can be found in A9 in APPENDIX A (see A10 for the English version).

Likewise, we were also interested in knowing the learners' impression of the task itself. The posttask questionnaire (DQ) consisted of 2 open-ended questions and 16 Likert-scale items (containing one negatively worded item). Among these items, 7 were related to common task aspects (such as perceived task difficulty or task repetition), whereas items 8-16 enquired learners specifically about the task setting in which they had completed the DG task (i.e. individual, collaborative or collaborative with a pretask FFI stage). The questionnaire can be found in APPENDIX A (see A11 and A12 for the Spanish and English version of the DQ completed by Comp and Collab; and see A13 and A14 for the Spanish and English version of the DQ completed by FFI+Collab).

When we piloted DQ, we noticed certain confusion with regards to the second block of Likert-scale items. All learners regardless of their set-up during the pilot DG intervention tended to answer the individual condition items, and others did not fill out the items concerning the collaborative condition despite having experienced that mode. As a result, there was a risk of having a high number of lost cases in those items. Therefore, in the final version we included colored boxes surrounding each of those item blocks to make the difference clearer to the children (a green box was used for the collaborative DG items and an orange box for the individual DG items). In addition, many learners found the Spanish term "tarea" (task) confusing, as it was mistaken for "homework" instead of being understood as the DG task. Thus, for the shake of clarity, some pictures of the original DG stories were included on the first page instructions.

The small sample number in the pilot test did not allow running an *a priori* factor analysis or any robust inferential statistical analyses. Therefore, the validation of both

AQ and DQ (by means of internal validity and subcomponent identification) was conducted once the final results were obtained and will be reported below, in the results of RQ3b.

5.4.2 Procedure

Before the data collection, the necessary approval was obtained from the Ethical Committee in Research on Human Beings, Samples and Data of the University of the Basque Country (see A15 in APPENDIX A). Having obtained the approval, the schools were contacted and a meeting was held with the English teachers and the children's class tutors in order to explain the experimental procedure. Some of the originally planned sessions had to be rescheduled and certain tests (as the attitude questionnaires and the writing tests) were concentrated on a single day, so as not disrupt the children's regular teaching too much. After the teachers agreed to participate, the two school boards and the children's parents granted permission to conduct the experiment (see the informed consent in A16). Since a few parents did not provide the necessary consent, we agreed with the teachers that some extra materials would be necessary for these children while the experimental sessions were being conducted. The experimental procedure followed in the present study is summarized in Table 6:

Table 6 The experimental design

Week 1	*Day 1*	AQ + L2 writing Pretest (T1)
	Day 2	L1 writing test
Week 2		Dictogloss (Day 1)
Week 3		Dictogloss (Day 2)
Week 4	*Day 1*	DQ + L2 writing Posttest (T2)
Week 5		—
Week 6	*Day 1*	Delayed L2 writing Posttest (T3)
	Day 2	Focus group interviews

In Week 1, child participants had approximately 30 minutes to complete the pretask attitude questionnaire (AQ) and 30 minutes for their L2 writing pretest. The L2 writing pretest will be henceforth referred to as T1. Similarly, learners had half an hour to complete the pretest story in their L1.

In Weeks 2 and 3, learners completed the two DG tasks (DG1 and DG2, explained above). Although the literature on TR does not specify the number of times a task should be performed or the intervals between each session, we decided to follow previous work on oral task repetition with EFL YLs. Hence, previous work with this population had already demonstrated changes in oral CAF with at least two repetitions (Azkarai & García Mayo, 2017; García Mayo et al., 2017; García Mayo & Imaz Agirre, 2016), and regarding intervals, a one-week interval has been most frequently used (Lázaro Ibarrola & Hidalgo, 2017; Pinter, 2007b, 2007a; Sample & Michel, 2015). Before learners performed the task for the first time, they were reminded that they had completed a similar exercise in their Basque lesson some weeks before (as explained above).

For the Collab and FFI+Collab groups, child learners were paired up in relation to the score they had obtained in the *Flyers* proficiency test (explained above, in *Participants and setting*). As could be observed in their results, Table 3 and Table 4 above, there was a great variability within each of the groups. Therefore, every effort was made so that there were not excessive proficiency differences between the members of each pair, as the CW literature had shown these to be detrimental for engaging in collaborative dialogue (Choi & Iwashita, 2016; Kim & McDonough, 2008; Kowal & Swain, 1994). On a 100-score scale, the difference between the pair members was the following: Collab $M = 7.86$, $SD = 6.25$, $Min = 1.2$, $Max = 23.48$; FFI+Collab $M = 6.2$, $SD = 6.61$, $Min = 0.17$, $Max = 19.11$. Therefore, we can confirm that the proficiency difference between the members of the same pair never exceeded 24 points and it was as low as almost 0.20.

Finally, in the case of the two participants who were assigned to the collaborative working mode (one from the Collab and another from FFI+Collab) who had not taken the *Flyers* proficiency test, their general English level was consulted with their English teacher, and assigned them to work with a partner of an estimated similar proficiency. With regards to FFI+Collab, due to the odd number of participants (n = 31), we decided to set a small group of three learners (MAR022, MAR027 and MAR035), as two previous study examining collaborative DG and EFL children had demonstrated that both conditions perform similarly and display analogous attitudes towards the task (Calzada & García Mayo, 2020a, 2020b).

The dictogloss procedure for the different experimental conditions and the comparison group was the following:

- Comp: those learners from School A who carried out the task individually stayed in their classroom with their English teacher. The English teacher, who had previously received the necessary instructions and the DG text recording, explained the goal of the task to the learners and played the audio twice. In the first listening, learners had to get a general idea of the text. During the second listening, the teacher handed them over the "Note-taking photocopy", where children could write key ideas or words from the text. The note-taking worksheets (see A3 in APPENDIX A) contained a picture related to the dictogloss passages, in order to provide children with a visual cue in case they found the recording too difficult. In DG1, there was a drawing of the main features of the story (the main character, some peanuts with a prohibition sign, and a teacher in a classroom), all taken from the open image repository *Openclipart*. In DG2, instead, there was a picture of a family in the garden which had been previously used in a POSS acquisition study with Basque-Spanish EFL learners (Imaz Agirre & García Mayo, 2013). In contrast to the original study, we decided to include the proper names of the main characters (María and Tom) and the family names of the rest of the characters in the story, in order to trigger the use of POSS in children's reconstructions.

  Next, they were provided with the reconstruction photocopy, and they were required to write a text that resembled as much as possible to one they had heard. They had 20 minutes as maximum to write the story. Finally, the teacher asked them to read the story they had written to themselves and gave them the "Written languaging" photocopy, where they could state the difficulties surrounding the text or the language. They could use as much time as they needed for this task.
- Collab: the pairs from School A who performed the DG task collaboratively were taken by a research assistant to a waiting room. Meanwhile, the main researcher and another research assistant called the different pairs one by

one to separate rooms, where a video camera and an audio recorder had been placed in order to record the learners' interaction. The first part of the task was the same as for Comp. That is, they listened to the audio twice and took down their notes individually. After that, they were encouraged to work together on the text reconstruction. They were given 25 minutes as maximum to do so. More time was allotted for this experimental group as it was considered that learners would spend some time on task management (for instance, deciding who was going to start writing or comparing their notes). They were given the choice to use the language they preferred for interaction and, if they happened to be speaking too low, they were kindly asked to speak up. To reduce the feeling of anxiety that the cameras or the recorder could produce, the researchers went around the room and did not observe them directly in order to create a sense of freedom. When the end time was approaching, the researchers reminded them that they only had a couple of minutes left. Finally, learners were asked to read the story aloud to check for possible mistakes.

- FFI+Collab: all participants from School B received some form-focused instruction (FFI) on the target forms (3S and POSS) right before engaging in the DG task. As the number of sessions had to be the same as in School A, for logistics reasons, the two 6th year primary classes from School B were split into two halves. One half (4-5 pairs) went to a separate room with the researcher in order to carry out the DG task, while the other half stayed in class with their English teacher. The following day, the latter swapped turns with the former. The room where learners had to carry out the DG task simultaneously was large enough so that the children would not disturb each other's interaction. Every table had a video camera and a voice recorder, to guarantee a correct recording quality. The FFI was delivered in a class format, that is, the teacher spoke to all four-five pairs at the same time. As explained in the "Instruments" section above, the FFI consisted of individual and pair work, as well as some researcher feedback. In order to make sure every student was correctly following the instructions, the researcher moved around the class and checked on every pair. After the FFI pretask stage, learners were introduced to the DG task. The procedure was the same as for the Collab group, and they were allotted 25 minutes as maximum to carry it out. The only difference was that before learners read their joint written product aloud, the researcher explicitly reminded them to bear the "golden rule" in mind. The golden rule referred to the target form linguistic rule which had been worked on during the pretask stage.

In Week 4, the same time allotment as in Week 1 was provided for the posttask questionnaire (DQ) and the L2 writing posttest (T2). Finally, in Week 6, children completed the L2 writing delayed posttest (T3) in thirty minutes, and one day after, several child participants (Comp n = 10; Collab n = 6; FFI + Collab n = 10) were recruited on a voluntary basis in order to take part in a guided focus group interview.

5.5 Codification and data analysis

5.5.1 LREs

The oral interaction from the collaborative DG (Collab and FFI+Collab), which totaled 13:43:11, was transcribed on Word following the CHILDES transcription conventions (MacWhinney, 2000). An example of a complete transcription can be found in A17 in APPENDIX A. Afterwards, LREs were identified and classified on NVivo (QSR International, 2012) according to the taxonomy used in previous CW research with YLs (Calzada & García Mayo, 2020a, 2021a; Luquin & García Mayo, 2021).

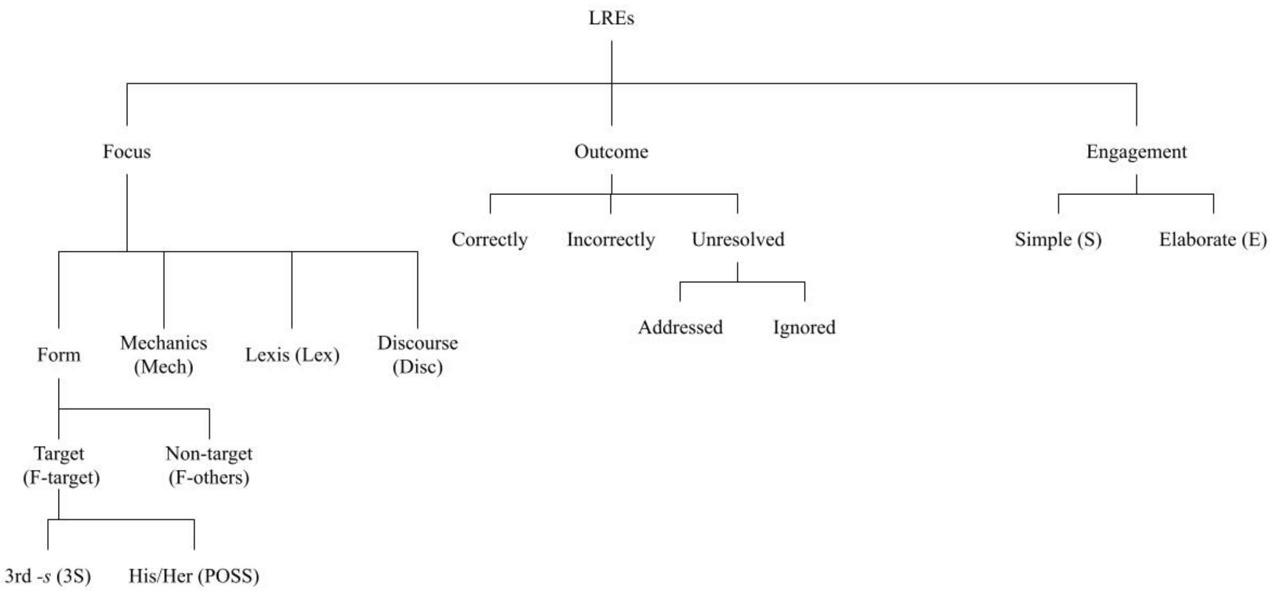
This classification is based on previous work with adult learners in ESL (Storch, 2008) and EFL settings (García Mayo & Azkarai, 2016), and it categorizes LREs regarding their (i) focus, (ii) outcome, and (iii) engagement. Nonetheless, it includes a more fine-grained description of the focus of children's discussions. Whereas traditionally LREs are split into two broad groups (meaning-based or form-based), we were interested in examining to what extent YLs attended to grammar, as opposed to other topics, namely, spelling and pronunciation, which have been generally included in "form". In fact, as previous studies have shown, child EFL learners devote much of their attention to these mechanical discussions (sometimes even constituting a greater proportion than grammar-related episodes, as shown in Calzada & García Mayo, 2020a, 2021a). Therefore, in order not to blur the results regarding the category of grammar, we labelled those related exclusively to the target forms (3S and POSS) "F-target", and those that dealt with other grammatical features (such as subject pronoun choice, article choice, infinitives and gerunds, etc.) were named "F-others". On the other hand, LREs revolving around spelling, punctuation and pronunciation were grouped into the category of mechanics (Mech).

In addition, LREs about lexis (Lex) were also separately tallied. Finally, children's discussions at a text level (including comments on the inclusion of a title, division of paragraphs or writing style) were clustered in the category of discourse (Disc). This last category is not as common as the aforementioned in the LRE literature (Fortune, 2005; Fortune & Thorp, 2001; Luquin & García Mayo, 2021), but since the present dissertation also looks into L2 writing development, we considered it relevant to provide information on discussions which could eventually impact on learners' L2 writing quality.

The second category, outcome, is related to children's ability to resolve LREs in a target-like manner. Hence, we considered "correctly resolved" those episodes in which participants arrived at an agreed outcome which coincided with the L2 norms, whilst the opposite was true for "incorrectly resolved" LREs. A third subcategory, "unresolved", referred to those discussions in which children were unable to find an agreed resolution. Moreover, it was further subdivided into two categories: "addressed" and "ignored". Addressed episodes were those in which, although both members of the dyad (or all three members of the small group) participated in the discussion, they eventually decided to move on to another topic as they lacked the certainty to provide a solution. Conversely, ignored episodes occurred when one member of the dyad (or the small group) indicated aloud either to themselves or to others a language-related problem and received no answer from their partner. It should be reminded that while both correctly and incorrectly resolved LREs were always reflected on the joint text, unresolved discussions were never translated into the co-constructed writing.

Finally, the third category (engagement) examined to what extent learners contributed to the discussion and showed positive signs of mutuality and equality. Although there have been different ways to operationalize how actively learners take part in metalinguistic discussions (Kuiken & Vedder, 2002a; Storch, 2008), we followed McDonough and Hernández-González's (2019) proposal, which classified engagement into elaborate (E) or simple (S). E engagement describes those situations in which "students worked together and each contributed to the resolution of the issue" (p. 111), whereas S engagement refers to those occasions on which only one member of the dyad provided the solution to a problem or the correction without an additional discussion. Figure 4 below summarizes the LRE classification:

Figure 4 Classification of LREs



In order to check for inter-rater reliability, following the procedure from other recent studies focusing on collaborative languaging (Baralt, 2014; McDonough & Hernández González, 2019), 20% of the oral dyad interaction dataset (that is, 12 randomly selected transcriptions) was independently coded by another researcher, who had previous experience in peer interaction analysis of SLA data. First, we wanted to determine the extent to which both raters agreed on what constituted an LRE. This is not a simple task given that it is a somewhat wide concept (Fortune & Thorp, 2001), which, as defined previously, includes dialogue as well as self-talk, and there might be instances where more than one LRE is encapsulated in a series of turns. This is the case of the following example:

(10) [MAR009 & MAR012 - Day 1]

*CHI B: Laura have to go to school but her father eh…
*CHI A: don't prepare…
*CHI B: **didn't? or forgot?** or but her father… → LRE about verb tense
*CHI A: but her father…
*CHI B: forgot.
*CHI A: don't or *¿cómo se llamaba?* (what was it called?)
*CHI B: *he dicho* 'forgot', *se ha olvidado* (I said 'forgot', he forgot)
*CHI A: **or forgots** prepare her sandwich. → LRE about 3S
*CHI B: [repeating while writing] her father…
*CHI A: fotarrr faterrr fotarrr.
*CHI B: **forgot**… **her lunch.**
*CHI A: **her sandwich!** *es que si no, no está bien* (otherwise it's not correct) [taking an object from the table] […]→ LRE about lexis
*CHI B: ***solo digo una cosa: si escribo mal sandwich es tu problema****.* (I'm only saying that if I misspell sandwich it's your fault)
*CHI A: sandwich.
*CHI B: ***no pero es que,* sand-wich*, ¿así?*** (no, but it's sand-with, like this?)
*CHI A: **no, no tiene hache** (it doesn't have an aitch) → LRE about mechanics (spelling)

In (10) we can see that both learners start deliberating about what tense they should use in the auxiliary ("don't or didn't"), but CHI A suddenly tries aloud to test whether the 3S sounds correct when added to the verb ("forgots"). Right after that, the topic shifts to a lexical choice between 'lunch' (CHI B's proposal) and 'sandwich' (CHI A's suggestion), and finally it is CHI A's word that they opt for, despite the difficulties they have in spelling it. As we can see, this collaborative dialogue illustrates the meticulous process for identifying LREs and setting their boundaries (one might even justify that CHI A is testing her own pronunciation of the word 'father' right between the lexical and the 3S-focused LREs). When learners deliberated different linguistic topics, even though they overlapped, these were coded as separate LREs (Nguyen & Newton, 2019). Other problematic cases were those instances in which learners discussed the same linguistic point several times, as in the following example:

(11) [ABE054 & ABE061 - Day 1]

*CHI A: some *cacahuetes*… (some… peanuts)
*CHI B: yes but I don't know.
*CHI A: buff… I don't know.
*CHI B: with eh…
*CHI A: *¿frutos secos?* (dry fruits?)

*CHI B:	I don't know.
*CHI A:	*a ver… frutos*, fruit eh… *seco…* (let's see… fruits, umm dry…)
*CHI B:	I don't remember, *bueno* (well)… umm with eh…
*CHI A:	fruits?
*CHI B:	with some… eh… *es que* (it's just that…) I don't remember the name.
*CHI A:	*¿'seco' cómo era?* (how was 'dry'?)
*CHI B:	*es que* (it's just that)… I don't know, we put then…
*CHI A:	*entre comillas*. (between inverted commas)
*CHI B:	or *espera* (wait) then here start the other…
*CHI A:	sentence.
	[…]
*CHI A:	because she forgot, she…
*CHI B:	she is, *bueno* (well), I don't know eh…
*CHI A:	is allergy…
*CHI B:	I don't know what is the… I don't know the…
*CHI A:	allergy… to…
*CHI B:	to this but I don't know the name [pointing at the Spanish word they used before] hahaha the same.

In (11), learners are struggling to find the English equivalent of *"cacahuetes"* (peanuts), one of the key elements in the DG1 story. They first have a discussion about it when they try to describe the chocolate that the character buys at the supermarket, and later on, they encounter the same problem when they want to explain that the character was allergic to this food. In this case, learners on both occasions were unable to find the target-like lexical solution. However, there were cases in which the second or third time they discussed the same linguistic unit learners eventually arrived at the correct resolution thanks, for example, to more engagement from both learners. Hence, we decided to count each instance of the same feature as an independent LRE.

To ensure some common understanding before the second coder conducted her analysis, some guidelines were provided (see A18 in APPENDIX A). After going through all the LREs identified by both coders and manually coding the agreement on an Excel spreadsheet, the percentage of agreement[14] between both raters was 60%, which was

[14] Regarding LRE identification, a percentage of agreement is provided instead of Kappa's coefficient because raters could only agree on what constituted an LRE but not on what did not constitute an LRE (i.e. the agreement count for the "Not LRE" condition would always be 0). Hence, the comparison did not meet the requirement to perform the inferential Kappa statistic, but it was performed for the subsequent analyses.

considered acceptable, taking account the large amount of LREs identified by both raters in the 12 dialogues (n = 200).

Once having ensured that the concept of LRE *per se* was reliable, another set of interrater analyses was conducted on the LREs identified by both raters (n = 116). In this case, a Kappa's correlation coefficient (κ) was calculated for each of the LRE classification categories: focus, outcome and engagement. Regarding focus, the Kappa value indicated strong and above chance inter-rate agreement: .89, CI [0.82 - 0.96]. In the other two dimensions the agreement was lower: in the case of outcome the value was .63, CI [0.51 - 0.75] and, with regards to engagement, κ was .60, CI [0.47 - 0.73].

In order to check for the main discrepancies in outcome and engagement, both coders went through the dataset again in order to determine the source of the differences in their criteria For instance, when assessing the type of resolution, the second rater had considered incorrect those cases where learners discussed a lexical choice, arrived at the target-like word but spelled it wrong in their writing. Those instances were, hence, counted as correct, since, as far as lexis was concerned, the dyad had resolved the LRE correctly. When analyzing the engagement of LREs, it was more complicated to discern what the main reason for disagreement was, as coders were dealing with a somewhat more subjective construct. However, it could be observed that the second coder had been more conservative when identifying "elaborate engagement" in LREs which were focused on mechanics (be it punctuation or spelling). The type of intervention regarding these two linguistic features was always short and never allowed for in-depth discussions, as compared to lexis or form. Yet, by comparing these LREs from the same pair and DG task, we could argue that two of them displayed more elaborate engagement than the rest:

(12) Elaborate engagement in punctuation-related LRE [ABE058 & ABE059 - Day 2]

*CHI B: *¿punto o coma?*
(full stop or comma?)
*CHI A: um comma, Maria says to her mum to her mother, Maria say to her mother to take some pictures and her father is laughing all the time, they laugh.

(13) Simple engagement in punctuation-related LRE [ABE058 & ABE059 - Day 2]

*CHI A: day, *punto.* (full stop)

(14) Elaborate engagement in spelling-related LRE [ABE058 & ABE059 - Day 2]

*CHI A: winning all the time.
*CHI B: *vale* (ok)
*CHI A: no, old, no [spelling in spa] a-l-l.
*CHI B: *¡ah, sí! ¡claro!* (oh yeah, of course!)

(15) Simple engagement in spelling-related LRE [ABE058 & ABE059 - Day 2]

*CHI A: [looking at the word] her, hum [meaning yes], uncle [pronouncing orthographically] un-cle to football, to the football, to football (…) to football.

Starting with examples (13) and (15), it can be observed that CHI A uses self-directed talk to refer to punctuation and spelling while she is writing, possibly as a way of confirming what she already knows in the L2. In both of these categories, an impressionistic analysis of the LREs led to the idea that self-directed talk was most frequent, that is, when the scribe made individual decisions aloud concerning spelling and punctuation. If these instances are juxtaposed with examples (12) and (14), a clear difference can be observed. In (12), CHI A consults her peer about the need to include a full stop or a comma at the end of the sentence, and CHI B answers what she considers most suitable (a comma). In (14), CHI A notices that the spelling of "old" is wrong, which leads CHI B to the notice her error and find the correct solution. Both (12) and (14) had been classified as simple engagement by the second coder, as they do not involve lengthy discussions about mechanics. However, a compromise was reached to classify as "elaborate" engagement those LREs about mechanics were both members of the dyad took part. In fact, in the case of low-proficiency child learners, we could not expect deliberations about mechanics involving more than a couple of turns.

5.5.2 Individual written languaging

For the written languaging generated by the learners in Comp, a similar procedure as with the oral LREs was followed. Rater 2 independently coded 50% of the written languaging dataset (n = 41). In this case, we only classified their written comments according to their focus category. Some of the children's comments had to be left out of the analysis as they did not match any of the linguistic categories, but they were rather metacognitive notes on the difficulty of the task or their ability to recall ideas. The extracts in (16) illustrate these cases:

(16)

No he entendido el audio (I didn't understand the audio) [ABE017 - Day 1]
No he entendido bien la lectura y lo he hecho lo mejor que he podido (I didn't understand the text and I did it as best as I could) [ABE025 - Day 1]
No me acordaba mucho de lo que decía el texto porque iba bastante rápido (I couldn't remember much from the text because it went quite fast) [ABE044 - Day 1]

Cohen's κ value for interrater agreement was .74 [CI 0.57-0.91], suggesting a strong level of agreement when assessing the topic of these languaging episodes.

5.5.3 Patterns of interaction

In order to ensure interrater reliability regarding pair dynamics, the second coder, who had previously assessed the LREs and the individual written languaging, went through the twelve interactions again. Examples of dyadic behaviors from other studies that had assessed the overall patterns were provided (Azkarai & Kopinska, 2020; Butler & Zeng, 2015). Given the small dataset, instead of calculating a correlation coefficient, both coders decided to review the interaction of those dyads where they had disagreed (5/12 dyads), and resolved those discrepancies orally. The features that should be present in the learners' interaction were clarified in order to classify their patterns as (i) collaborative, (ii) dominant/dominant, (iii) passive/parallel or (iv) dominant/passive. Afterwards, the first coder reviewed the complete dataset (29 dyads) and introduced some changes where necessary.

5.5.4 L2 written product analysis

The handwritten L2 written production from (i) the pretest, posttest and delayed posttest and (ii) the individual or collaborative DG tasks (see A19 in APPENDIX A for an example of a handwritten text) was transcribed on Word and analyzed using text-based measures as well as overall quality measures from an analytic rubric. As the literature shows, it is important to use both kind of measures because one cannot always associate specific text measures with increased quality (Polio & Friedman, 2016, p. 119). What is more, although TBLT research has extensively used CAF text measures to assess task-based performance, rubrics are more ecologically valid as the constitute the main means for evaluating writing in school contexts (Polio & Shea, 2014).

5.5.4.1 Text-based measures

A range of manual and automatic procedures were used to evaluate the written production. The majority of measures (the clause type classification, the grammatical complexity measures, as well as the general and specific accuracy measures) were coded manually on an Excel spreadsheet divided by sentences and rows with measure frequencies (Polio & Friedman, 2016, p. 117).

After reviewing the previous literature, we decided to concentrate on the analysis of grammatical and lexical complexity, as well as accuracy, leaving out the dimension of fluency. In fact, in the field of writing, fluency is still not exempt of controversy (Elabdali, 2021; Michel, 2017; Polio & Friedman, 2016), as some frequency measures that have traditionally been employed as fluency indices (e.g. total number of words, clauses/T-unit) are now considered indicators of complexity (Norris & Ortega, 2009), especially in those cases where the time spent on the task is not kept constant.

In this vein, writing offers opportunities for reflection, monitoring and editing, and hence, certain scholars working in the field of TBLT question the necessity to quantify a measure which has been directly borrowed from orality, where speaking faster does correlate with a more target-like performance (Ahmadian & Mansouri, 2020). In our context, where child learners were not encouraged to write faster and had a relatively flexible time span to produce their texts (as explained before), fluency was not considered a relevant construct. Finally, in order to determine how well child learners had fulfilled the DG task main requirement (i.e. to reconstruct the original story as faithfully as possible), a measure of content accuracy was included within the specific accuracy dimension. Table 7 summarizes the text-based measures:

Table 7 Summary of text-based measures

Complexity	Clause type based on grammatical and functional complexity (Torras, 2005; Torras et al., 2006)	Preclause ratio (PreCRatio)
		Protoclause Ratio (ProtoCRatio)
		Paraclause ratio (ParaCRatio)
		Full-fledged clause Ratio[15] (FullCRatio)
	Interdependence grammar complexity	Clauses per T-unit (C/T-unit)
		Coordinated clauses per T-unit (Coord/T-unit)
	Length-based grammar complexity	Mean length of clause (MLC)
	Lexical complexity	Guiraud's index (GI)
Accuracy	General	Grammar errors per 100 words (GramErr100)
		Grammar error Free Clauses (GramErrFreeC)
		Spelling errors per 100 words (SpellErr100)
		Borrowings per 100 words (Borrow100)
	Specific	Target form accuracy (3SAcc & PossAcc)
		Content accuracy - Idea Units (IUs)

These were the criteria for each of the manually coded measures (all the examples are taken from the current dissertation data and errors are kept as in children's original texts):

[15] In Torras' (2005) original classification these were labelled simply 'clauses', but we added 'full-fledged' in order to distinguish them from the pure syntactic criteria used to tally clauses.

Complexity

- Grammatical and functional complexity: we were guided by Torras' (2005, pp. 96-97) definition of the following units, which take the finite clause as the central element to understand grammaticality, while the writing task requirements determine whether clauses are contextualized or not:
  - Preclause: these units to do not meet the requirement of either grammaticality or contextualization. They are produced by the learner to show some knowledge of the target language and lack the necessary components to be considered clauses. They also contain numerous morphosyntactic, lexical and spelling errors: "and hospital can ales in potato" (ABE021 - Day 1); "and hits "*cacahuetes*" [peanuts] and "*tuvo*" *barisels* [got chicken pox]" (MAR022 & MAR027 & MAR035 - Day 2).
  - Protoclause: the content of the clause is related to the task requirement but it does not meet the criteria to be considered grammatical: "Nora is to go is hospital" (MAR003 & MAR007 - Day 1); "Maria they play with grandfather" (ABE008 & ABE013 - Day 2).
  - Paraclause: although the clause is grammatical, the content is completely unrelated to the task, more likely written to lengthen the composition: "Laura is one person" (ABE048 - Day 1); "He's a doctor" (ABE032 & ABE036 - Day 2).
  - Full-fledged clause: the clause is both contextualized and grammatical, although there may be minor morphosyntactic, lexical and spelling deviations: "Lora have to go to shcool, but her fater forgot her sawich" (MAR009 & MAR012 - Day 1); "One day the Smith were on the garden, Tom and his uncle were playing football" (ABE003 & ABE010 - Day 2).

The ratios of each clause type (PreCRatio, ProtoCRatio, ParaCRatio and FullCRatio) were obtained dividing the number of clauses from a specific category by the total number of clauses.

- *Structural complexity*
    - Clauses: following the linguistic definition of clause strictly (Bulté & Housen, 2014), these were considered units consisting of a subject (explicit or implied) and a predicate, which combined together form sentences. Consequently, verbless constructions, also known as subclauses in oral CAF literature (Foster et al., 2000), were not considered clauses. Furthermore, in contrast to Bulté and Housen (2014), verb clusters such as "goes to buy", "tries looking for" or "starts to play" were analyzed as a single main clause, since counting them as separate clauses can distort the calculation of clausal complexity measures, such as MLC (Bulté & Housen, 2012; Mylläri, 2020). Although sentences can consist of a simple clause, they can be made longer and more complex by means of coordination and subordination:
        - Coordinated clauses: these are syntactically homogeneous clauses connected through coordinating conjunctions (and, but, so, etc.)
        - Subordinated clauses: this type of dependent clause is introduced in a main clause by means of subordinating conjunction (when, where, that…). They minimally consist of a finite or non-finite verb plus at least one other element (e.g. a subject or an object) (R. Ellis & Barkhuizen, 2005, p. 148).

    Examples (17) and (18) show how child learners' texts were coded for different types of clauses. The symbol (/) limits simple independent clauses, (//) establishes the boundary between coordinated clauses, and finally, ([]) is used to mark dependent subordinate clauses.

(17)

The smiths are spending the day on the garden./ Tom and his uncle are playing football they call grandmother to enjoy it, // but she doesn't look [because is playing cards with the grandfather // and she is winning all the time]. Maria says to shes mother to take some photos // and dad is puting funny faces to laugh./ Is a very hot afternoon./ Tom says [his grandfather is preparing ice cream in the kitchen]./ Maria goes very kuickly to taste / it is delicous./ Maria takes a porcion // and brings it to the aunt, [is in the swiming pool]. What a beautiful day!
[MAR025 & MAR029 - Day 1]
5 simple independent clauses, 8 coordinated clauses, 3 dependent clauses

(18)

Lora fader forgot put the sanwich // and give mony to pay the sanwich // and go to supermarquet se chocolate in the supermarquet // and go to pay the sanwhit / Lora goes to scool / lora is alerllic to the pinuts // and lora go to Hospital // and go her hose...
[ABE062 - Day 1]
1 simple independent clause, 7 coordinated clauses

Although sentences are employed in a large number of studies assessing written grammatical complexity (Bulté & Housen, 2014; Lahuerta Martínez, 2017; Vyatkina, 2012), they were not deemed informative in our context due to children's inappropriate use of punctuation marks (Mylläri, 2020), which usually led to extremely long stretches of text between two full stops. In contrast, the concept of T-unit, that is, an independent clause with all its dependent clauses (Hunt, 1965), which is also common in the young EFL writing literature (Hidalgo & Lázaro Ibarrola, 2020), had the potential to provide more information about the interrelation type between clauses. The number of T-units was calculated by adding the two types of independent clauses (simple and coordinated), and afterwards, the following ratio measures were obtained:

- Clauses per T-unit (C/T-unit): the number of clauses per T-unit has been shown to be a clear indication of the amount of clausal subordination (Kyle & Crossley, 2018; Lu, 2011).

   - Coordinated clauses per T-unit (Coord/T-unit): Coordination has been claimed to be an indicator of complexity at beginning stages of proficiency (Norris & Ortega, 2009; Torras et al., 2006; Vyatkina, 2012). In order to calculate this ratio, the number of coordinated clauses was divided by the total T-units.
   - MLC: the total number of words (tokens) was divided by the total number of clauses

- *Lexical complexity*

  In order to gauge quantitatively the construct of lexical complexity we used Guiraud's Index (GI) for lexical diversity or richness (Guiraud, 1959). This index belongs to the family of type/token based ratios, with the advantage that it includes a correction for text length, so that not only texts with fewer lexical repetitions are considered "richer", but even more so those that are longer (Bulté & Housen, 2014). Conversely, the short length of the child learners' texts in the present study (the overall average number of words was M = 55.34, SD = 35.83) did not allow a widespread use of D (a more sophisticated statistic also based on a type/token ratio), as it requires texts to be above 50 words (Meara & Miralpeix, 2017).

  GI is obtained by dividing the number of types by the square root of tokens. Although the actual number of types and tokens was automatically obtained using *Voyant Tools* online software (Sinclair & Geoffrey, 2016), some manual handling was needed prior to that step. First, we needed to determine what unit we were going to base our analysis of types on. As shown by the literature (Schmidt, 2010), lemmas seem to be a reasonable unit for learners who already master the inflectional suffixes (such as the plural -*s* marker). Moreover, evidence from the field of psycholinguistics (Aitchison, 2003) has shown that learners store lexis on a lemma-based fashion. Therefore, as an example, 'picture' and 'pictures' were counted as a single type (but see Silva & Clahsen, 2008 for counterevidence regarding complex derivational morphology).

Besides, contractions of the verb 'to be' (e.g. 'it's') were treated as single words. Learners frequently misspelled contractions, suggesting that children of this proficiency may not probably have assimilated the morphosyntax behind them: "In the afternoon its bery hot." [ABE041 - Day 1].

Likewise, spelling mistakes were corrected as long as the words were clear enough to convey the meaning. It is important to edit these mistakes, as any software would count 'school' and 'shool' as two different types, to the detriment of learners who not only were more diverse in their writing, but were also better at spelling. In the same vein, lexical errors were taken out of the analysis, as they did not provide evidence of the learners' productive vocabulary (Pawlak, 2016). Finally, invented L2 words (*she's father → her father) and L1 words and expressions were also left out of the GI.

Accuracy

- *General measures*
  - Grammar errors per 100 words (GramErr100): any deviations from target-like grammar use except for the errors related to the use of 3S and POSS (as these were included within specific measures) were considered grammar errors. The calculation was done dividing the total number of errors by the total number of words (tokens) produced divided by 100 (R. Ellis & Barkhuizen, 2005).
  - Percentage of grammar error-free clauses (GramErrFreeC): bearing in mind the concept of grammar clause explained above, we divided the number of error-free clauses by the total amount of clauses divided by 100 (R. Ellis & Barkhuizen, 2005).
  - Spelling errors per 100 words (SpellErr100): we calculated the number of spelling errors divided by the total number of words (tokens) produced divided by 100.

- L1 transfer lexical errors per 100 words (Borrow100): this type of lexical errors typically include borrowings and coinages (Agustín Llach, 2011, p. 123)., and have been found most frequent in low proficiency YLs' writing (Navés et al., 2005). Borrowings are also called "code switching" and are more recently referred to as translanguaging. These terms describe the situation when the learner inserts any L1 word into the L2 syntax without making an effort to adapt it to the L2 morphology or phonology. Here are two examples: "I go to supermarcket because shes sale a *aguacate*" (sell an avocado) [ABE046 - Day 1], "Uncle as "*Ikusi nola jolasten dute kartasera*"" (Uncle sees how they play cards) [MAR032 & MAR034 - Day 1]. Conversely, when learners make use of coinage, they attempt to adapt the L1 word to the conventions of the L2 orthography and morphology: "is Family her in the meriend and tom is playing fotball" (**meriend* -> *merienda,* snack) [MAR023 & MAR033 - Day 1]. In order to calculate the lexical error ratio, the amount of borrowings and coinages was divided by the total amount of words (tokens) divided by 100.

- *Specific measures*
  - Target form accuracy: The procedure for calculating accuracy for the two target forms was based on Pica's (1983) "target-like analysis" formula, which divides the *n* correct suppliance in obligatory contexts by all obligatory contexts plus *n* suppliance in non-obligatory contexts. In what follows, our understanding of obligatory context for each of the target forms is defined.
    - 3S accuracy rate: wherever the learner used the present tense with a third person singular subject was considered an obligatory context. If the learner, instead, clearly used the past tense although the original story was set in the present, these instances were not considered obligatory 3S contexts. Overproduction of 3S occurred, for example, when the learner attached the morpheme to the verb form

with a subject pronoun other than third person ("In the night they **goes* to ours houses", [ABE028 - Day 1].

- POSS accuracy rate: all family names where the story provided clear possessor references (grandmother, grandfather, aunt, etc.) were considered obligatory contexts. If the learner used a zero article or the definite article, this was considered a non-target like use ("And Mother is taken fotos", [MAR001 & MAR014 - Day 2]; "Then, in the class **the* teacher see Lora's faces", [ABE030 & ABE042 Day - 2]. Personal belongings and body parts were also considered obligatory contexts whenever there was a preceding possessor. Overproduction occurred when the learner, for instance, used POSS instead of the subject pronoun ("and **her* say to teacher then teacher col to Loras fater and Lora go to hospital", [ABE023 - Day 2]). The use of the Saxon genitive to avoid employing POSS was not taken into account ("The teacher calls <u>Noras dad</u> and he calls the doctor", [MAR030 & MAR031 - Day 2].

- Content accuracy (CA): in addition, in the case of the texts learners wrote as a response to the DG task, we aimed to determine the degree of task fulfilment by counting the number of Idea Units (IUs) (Carrell, 1985). The original texts were divided into 14 Idea Units each. To help the coder decide, an abstraction of the idea units was added (see A20 in APPENDIX A).

As in the case of the LRE and dyadic pattern codification, all text-based measures for L2 writing (except for lexical complexity, which was automatically obtained by means of a software) were coded by two independent raters. In order to do so, 20% of the written production was selected for interrater analysis. Pearson correlations were run for the following units of analysis: preclause (r = .37), protoclause (r = .24), paraclause (r = .57), full-fledged clauses (r = .90), simple independent clauses (r = .88), coordinated clauses (*r* = .90), dependent clauses (*r* = .74), grammatical errors (*r* = .75), grammar error-free clauses (*r* = .64), spelling errors (*r* = .80), borrowings (*r* = .42), 3S and POSS obligatory contexts (*r* = .63; *r* = .71), 3S and POSS correct instances (*r* = .94; *r* = .67), and 3S and POSS overproduction instances (*r* = .92; NA[16]).

As expected, the measures with the lowest correlation coefficients corresponded to those categories related to the contextualization and grammaticality of the clauses, two constructs which are only vaguely described by Torras (2005). The two coders deliberated together all the instances where they had disagreed, and found that the main source of divergence was the threshold of tolerance towards grammatical deviations in these clauses, which on most occasions were hindering their comprehensibility (hence, blurring the difference between preclauses and protoclauses). After this discussion, only those clauses that were totally unrelated to the original text content, contained three or more errors or did not have a finite or non-finite verb as nucleus were considered preclauses. Instead, if there were more than three grammar errors but the clause still conveyed some text-related information, it was considered an example of protoclause. Regarding the disagreement in the category of lexical errors, we observed that it could have been motivated by a lack of precision in the interrater guidelines where this measure was explained. In fact, the scenario were learners make use of coinage had not been included, which led the second coder to identify fewer errors in this category. Once again, the two coders went through the selected dataset again until 100% agreement was reached.

With regards to IUs, since it was a categorical variable, the Kappa correlation coefficient was calculated for the fourteen units in DG1 and DG2. The results showed a satisfactory reliability in most IUs from both texts (κ > .70), but there were certain exceptions. The coders analyzed jointly the problems of those IUs. In DG1, for instance, for IUs 5, 7 and 8 a low coefficient was found (κ =.41 - .42). IU 5 could be easily confused with the content of IU 4 and IU 6 and, therefore, we decided to emphasize the action of "seeing" in IU 5. Conversely, IUs 7 and 8 were found to contain more elements than the rest of the units, and hence, their abstraction was simplified so that they only referred to the key elements for keeping the sense of the story ("buying peanuts" and "eating",

[16] None of the coders identified instances of POSS overproduction in the dataset explored for this analysis.

respectively). Regarding DG2, the ideas with the lowest reliability were IU 2, 5 and 10 (κ = .32 - .44). These abstractions were modified so that they could better capture the main idea. More specifically, IU 2 was extended from "football" to "Tom uncle football", as all three elements were considered relevant. In IU 5, "sitting at the table" was chosen as an abstraction instead of "table garden", and finally, in IU 10 "trying ice cream" was reformulated as "having ice cream", to prevent assessing the whole unit too strictly due to a different lexical choice made by the child learners.

5.5.4.2 Rubric assessment

The rubric employed in the present dissertation (see A21 in APPENDIX A) is based on the one designed by the Department of Education of Navarre (Government of Navarre, 2017), and had been used in previous research on CW with EFL adolescents in the first years of secondary school (12-13 years) (Villarreal & Munarriz-Ibarrola, 2021). Each piece of writing was assessed on a 1-3 scale (1 being the lowest and 3 the highest) for these categories: adequacy (Adq), coherence (Coher), cohesion (Cohes), grammatical accuracy (Acc), mechanics (Mech) and lexis (Lex). For each category, every scale level contained a descriptor with the main achievement indicators that the evaluator should bear in mind when assessing the texts. In this sense, the wording of category of "adequacy" was slightly changed to assess the texts produced as a result of the DG task. In this case, this dimension explicitly referred to the retrieval of ideas from the original text. The rest of the rubric was kept identical (see A22 in APPENDIX A). Spearman (ρ) correlation analyses were run in order to check for interrater agreement. The results indicated a general high correlation between the two coders for adequacy, coherence and cohesion ($\rho > 0.7$). However, in the case of accuracy, mechanics and lexis, the rates were slightly lower (from .50 to .58). The two coders resolved their discrepancies for each case until a 100% agreement was reached.

5.5.5 L1 written product assessment

For the assessment of the L1 written production, an adapted version of the rubric designed by Fernández et al. (2019) was used, known as *Rúbrica de evaluación de relatos* (Story Assessment Rubric). This tool was originally developed for evaluating the narrative competence in the intermediate stage of compulsory schooling in Spain (ages 8-14), and was validated by means of expert evaluation (content validity) and confirmatory factor analysis (structural validity). It consists of seven dimensions (frame, topic, plot, creativity and interest, sentences, vocabulary and spelling), which are rated on a 1 (lowest) to 4 (highest) scale. The levels of each dimension contain a brief descriptor to guide the rater's decisions. In the present dissertation, the criterium of "frame" was left out of the rubric, as the introduction was already provided in the continuation task (explained above). The rest of the rubric was kept as in the original English version designed by Fernández et al. (see A23 in APPENDIX A). As in the case of the L2 writing rubric, this tool was also checked for interrater reliability. The second rater was asked to evaluate 50% of the sample ($n = 46$), and afterwards, Spearman correlation analyses were run on SPSS 24 (IBM Corp, 2016): topic $\rho = .36$; plot $\rho = .57$; creativity $\rho = .54$; sentences $\rho = .41$; vocabulary = .25; spelling = .49. As these values were far from an acceptable correlation rate ($\rho > .70$), the two raters went through the dataset again and resolved their discrepancies orally. In order to check again for the interrater reliability, the second half of the sample ($n = 47$) was evaluated by the second rater. The Spearman correlation analyses on this occasion were all above .70.

CHAPTER 6 RESULTS

In this chapter, the results in response to each set of RQs will be presented: RQ1 Dictogloss, Focus on Form (FonF) and the impact of task-related variables; RQ2 Collaborative work; and RQ3 Individual differences.

6.1 RQ1 Dictogloss, Focus on Form (FonF) and the impact of task-related variables

a) What is the focus of YLs' written languaging (wLREs) and oral language-related episodes (LREs)?
b) What is the LRE outcome, depth of engagement and Previously Known Language (PKL) use?
c) What is the impact of procedural task repetition (TR) and pretask focus on form instruction (FFI) on YLs' focus on form?
d) Do YLs focus on the two target forms (3S and POSS)?

In order to answer the research questions related to this topic, in RQ1a and RQ1b the results obtained from the analysis of the languaging (written and oral) on Day 1 will be first presented separate from the results obtained on Day 2. This will help to discern the effect of the task on languaging from the impact of TR, which will be shown in the comparison between Day 1 and Day 2 (in RQ1c). As all participants completed DG1 (focused on 3S) and DG2 (focused on POSS) randomly on Day 1 and Day 2, the turns and LREs of these two features are displayed together (as F-targets) in the analysis of the results of the first three questions, whereas the amount of focus on form devoted to each of the forms (regardless of testing day) is presented in RQ1d.

First, the results of the written languaging by the Comp group on Day 1 are reported. On this day, four learners failed to attend school and, therefore, there were thirty participants in total. Their reflections were classified in terms of focus, as displayed in Table 8:

Table 8 Written individual LREs in Comp on Day 1 / focus

	wLREs (n = 30)			
	N (%)	**M**	**SD**	**95% CI**
F-targets	0	0	0	--
F-others	2 (5.40)	0.07	0.25	[-0.28 - 0.16]
Lex	18 (48.60)	0.60	0.77	[0.31 - 0.89]
Mech	15 (40.60)	0.50	0.73	[0.23 - 0.77]
Disc	2 (5.40)	0.07	0.25	[-0.03 - 0.16]
Total	37	1.23	1.25	[0.76 - 1.70]

Table 9 Paired-samples t-tests for wLREs on Day 1 / focus

wLREs (n = 30)				
Comparison	**t**	**df**	***p***	**Cohen's *d***
F-others - Lex	-3.76	29	< .001	-0.69
F-others - Mech	-3.50	29	.002	-0.64
F-others - Disc	--	29	--	--
Lex - Mech	0.55	29	.586	0.10
Lex - Disc	3.56	29	.001	0.65
Mech - Disc	3.07	29	.005	0.56

The results of the written languaging from Day 1 show that children produced a small amount of metalinguistic reflections, and that most were related to lexical and mechanical aspects (adding up to 89.2% of all comments). Moreover, as shown by the statistical analysis in Table 9, the difference between these two aspects and F-others and Disc reached significance ($p < .05$), although the magnitude of those differences was small (d close to 0.60)[17]. Children expressed their doubts and difficulties in the following terms: *"no sé cómo se dice 'cacahuete' en inglés"* (I don't know how to say 'peanut' in English) [ABE014 - Day1], *"he escrito 'beby' así pero no estoy seguro de que esté bien"* (I have written 'beby' like this but I'm not sure if it's correct) [ABE024 - Day1]; *"he escrito 'weather' pero no se se escribía 'wether'* (I have written 'weather' but I'm not sure if it was 'wether') [ABE044 - Day1]. Examples of the few languaging episodes related to Disc and F-others (constituting roughly 11% of all wLREs) are also provided: *"¿cómo se dice 'have' en pasado?"* (how do you 'have' in the past?), *"no estoy segura de que esté bien*

[17] For the interpretation of Cohen's d, we were guided by Plonsky and Oswald's (2014) benchmarks. For within-group contrasts: small (d = 0.60), medium (d = 1) and large (d = 1.4).

escrito, no se me da muy bien escribir en inglés" (I'm not sure whether it's correctly written, I'm not good at writing in English) [ABE001 - Day1].

Finally, no child in the Comp group reflected on the use of the target forms. Likewise, the low mean values are worth noting, indicating that a large proportion of children (44%) did not produce any wLREs and could only comment on the difficulty of the task or the comprehension of the passage. Moreover, it should be highlighted that all wLREs were based on the sentence structures provided in the Written Languaging Worksheet, and these children did not use their own words to express doubts or difficulties.

Now the results of the oral languaging (LREs) generated by the learners who carried out the DG task collaboratively (Collab and FFI+Collab) are presented. For the first two questions (RQ1a and RQ1b), the results of both experimental groups are amalgamated, and later (in RQ1c and RQ1d), these will be displayed separately to account for the possible impact of the pretask FFI on the attention to form and other performance measurements (i.e., time on task, resolution and type of engagement in LREs and PKL use). Table 10 contains the descriptive statistics for LREs and the number of turns in each of the topic categories.

As far as LREs are concerned, the raw proportions indicate that dyads devoted most of their attention to mechanics and grammar (36.2% and 32.5%, if F-targets and F-others are counted together). These two formal aspects of language constitute almost 70% of all episodes, followed by Lex, and far behind by Disc. The statistical analysis (displayed in Table 12) revealed that, on average, dyads generated significantly more episodes about Mech than about F-targets, F-others, Lex and Disc, with effect sizes ranging from small (Mech vs Lex, Mech vs F-others) to large (Mech vs F-targets, Mech vs Disc). In addition, learners also produced significantly more episodes about F-others than about F-targets and Disc (with medium effect sizes). Finally, they also produced significantly more LREs about Lex than about F-targets and Disc (with medium effect sizes).

However, regarding the number of turns of these episodes (also displayed in Table 10), there are some clear differences. Additionally, in order to see the relationship between LREs and turns, we also calculated the average length of LREs, dividing the total amount of turns by the total amount of LREs in each topic category (turn/LRE). The results for this measure are shown in Table 11. The most numerous and lengthiest turns that children generated were those related to Lex (32%, averaging 3.52 turns/LRE). In fact, the statistical analysis (Table 13) showed that dyads produced significantly more turns about Lex than about F-targets and Disc (with medium effect sizes). In the case of Mech, the proportion of turns reflects a decrease if compared to the proportion of LREs in the same category (36.2% of all LREs vs 28.7% of all turns), indicating that children, on average, employed 2.18 turns/LRE to discuss spelling, punctuation or pronunciation. However, the difference in the average of Mech turns and F-targets and Disc turns also reached significance (with medium effect sizes). Conversely, the percentage of F-others turns (25%) represents roughly the same proportion as the LREs (26%), which translates into a slighter higher average LRE length than Mech (2.47 turns/LRE). The difference in the average between F-others turns and F-targets and Disc turns was also significant (with low and medium effect sizes, respectively). Finally, children produced the lowest number of turns related to F-targets and Disc (equaling approximately the same proportion as LREs), with a mean length which did not exceed two turns per LRE (1.44 and 1.28, respectively).

Regarding the target linguistic features specifically, as can be inferred from the raw proportion of LREs and turns (6.3% and 9%, respectively), learners who worked in the DG task in pairs did not devote a large amount of discussions to 3S or POSS, and when they did so, they did not discuss these features in length. Apart from that, as in the case of wLREs, the dispersion values (SD) also suggest a large variability among dyads both with regards to LREs and turns (which will be later scrutinized in terms of instruction type). This variability is especially evident in the case of F-targets, F-others and Disc, where the SD value comes close or even exceeds the mean value.

Table 10 LREs and turns / focus

	LREs (n= 29)				Turns (n = 29)			
	N (%)	**M**	**SD**	**95% CI**	**N (%)**	**M**	**SD**	**95% CI**
F-targets	33 (6.3)	1.14	2.06	[0.35-1.92]	137 (9)	4.72	6.97	[2.07-7.38]
F-others	136 (26.2)	4.69	3.29	[3.43-5.94]	385 (25.3)	13.27	12.83	[8.39-18.16]
Lex	135 (26)	4.65	2.89	[3.55-5.76]	497 (32.7)	17.14	12.79	[12.27-22]
Mech	188 (36.2)	6.48	3.41	[5.19-7.78]	437 (28.7)	15.07	10.74	[10.98-19.15]
Disc	27 (5.2)	0.93	1.41	[0.39-1.47]	65 (4.3)	2.24	3.21	[1.02-3.46]
Total	519	17.89	9.04	[14.46-21.33]	1521 3553*	52.45 122.52	26.43 53.06	[42.39-65.50] [102.34-142.70]

Note: Total turns (with and without LREs)

Table 11 Mean length of LREs 1 / focus

	LRE length (n = 29)		
	M	**SD**	**95% CI**
F-targets	1.44	3.04	[0.28-2.60]
F-others	2.47	1.21	[2.01-2.93]
Lex	3.52	1.66	[2.89-4.15]
Mech	2.18	0.67	[1.93-2.44]
Disc	1.28	1.59	[0.67-1.88]
Total	2.94	0.90	[2.59-3.28]

Table 12 T-tests for LREs on Day 1 in all pairs / focus

Comparison	t	df	*p*	Cohen's *d*
F-targets - F-total	-5.355	28	< .001	-0.99
F-targets - Lex	-6.086	28	< .001	-1.13
F-targets - Mech	-10.447	28	< .001	-1.94
F-targets - Disc	0.459	28	.650	0.08
F-other - Lex	0.057	28	.955	0.01
F-other - Mech	-2.533	28	.017	-0.47
F-other - Disc	6.319	28	< .001	1.17
Lex - Mech	-2.722	28	.011	-0.50
Lex - Disc	6.853	28	< .001	1.27
Mech - Disc	9.026	28	< .001	1.68

Table 13 T-tests for turns on Day 1 in all pairs / focus

Comparison	t	df	*p*	Cohen's *d*
F-targets - F-total	-0.545	28	.007	-0.54
F-targets - Lex	-0.823	28	< .001	-0.82
F-targets - Mech	-0.874	28	< .001	-0.87
F-targets - Disc	0.325	28	.091	0.32
F-other - Lex	-0.237	28	.213	-0.24
F-other - Mech	0.897	28	< .001	0.90
F-other - Disc	0.141	28	.453	0.14
Lex - Mech	1.115	28	< .001	1.11
Lex - Disc	1.236	28	< .001	1.24

To conclude RQ1a, examples for each of the LRE topics are provided and commented below:

(19) F-targets

(19a) [ABE007 & ABE011 - Day 1]

*CHI B: to go to school but her father forgot prepare her sandwich.
*CHI A: 'his' *sería mejor* (would be better), "his sandwich".

(19b) [MAR004 & MAR019 - Day 1]

*CHI B: Lola I love chocolate, Lola loves chocolate prr [as if she is not sure].

In (19) we can find an episode related to POSS (19a) and another to 3S (19b). In the former, CHI A amends CHI B's use of 'her', although the correction is not target-like (the reference possessor in the original story was a female character). In the latter, CHI B reflects on the use of the morpheme by means of self-directed talk, but she cannot decide on which option ('I love' or 'loves') is more appropriate.

(20) F-others

(20a) [ABE052 & ABE064 - Day 1]

*CHI A: after they go…
*CHI B: they go… to the hospital.
*CHI A: no, ↑she go to the hospital.

(20b) [MAR030 & MAR031 - Day 1]

*CHI B: the celeb a celeb… *a ver* (let's see) ¿the celebration *o* (or) a celebration?
*CHI A: a celebration.
*CHI B: a celebration.

Moving now onto the next LRE focus (F-others), example (20a) illustrates an episode related to the use of personal subject pronouns: CHI A uses 'they' when he actually refers to a female character, and right after CHI B's intervention, he notices this error and corrects himself. However, in (20b), CHI B enquires her partner about the article use (‘a’ vs ‘the’), and CHI A provides the correct alternative, which is then echoed by CHI B.

(21) Lex

(21a) [ABE033 & ABE037 - Day 1]

*CHI B: his make photos.
*CHI A: take!
*CHI B: *o sea* (I mean), yes take.
*CHI A: *joe…* (c’mon…)

(21b) [MAR006 & MAR018 - Day 1]

*CHI B: her lunch?
*CHI A: box, lunch box.
*CHI B: *¿libro o qué? ¿qué es eso?* (a book? Or what is that?)
*CHI A: *eh… es que no sé cómo se llama, donde traes la comida que es como una caja la merie… el sándwich y todo.*
(erm… I don’t know what you call it, where you bring your food, like a box, the snack… the sandwich and everything)
*CHI B: *¿el túper?* (the tupper?)
*CHI A: *sí, algo así.* (yeah, something like that)

With regards to Lex, in (21a) CHI A corrects a collocation (‘take photos’), which had been wrongly used by CHI B. CHI B is made aware of this mistake and repeats the correct collocation. On the other hand, in (21b), CHI B does not seem to know the meaning of ‘lunch box’ (which he mistakes for ‘books’) and CHI A explains in her own words in the L1 what it refers to, eventually finding an equivalent in his repertoire (‘tupper’).

(22) Mech

(22a) [ABE030 & ABE042 - Day 1]

*CHI B: and h... she’s winning, *como que va ganando* (like she’s winning), and…
*CHI A: *entonces en vez de punto ponemos* (so then instead of full stop we put) “and is winning”, “is playing cards and is winning”.
*CHI B: *con doble ene* (with double en).
*CHI A: *ah claro, porque es vocal consonante vocal, no consonante vocal consonante, eso,* winning umm and is winning, *punto*.
(oh sure because it’s vowel consonant vowel, not consonant vowel consonant that’s it winning umm and it’s winning full stop)

(22b) [MAR001 & MAR014 - Day 1]

*CHI A: today Lola… to go to school… go to the school.
*CHI B: *punto* (full stop).
*CHI A: and her parent, *punto* (full stop), prepared sandwich, *no hay punto ahí* (there's no full stop there), and, and…
*CHI B: *no sé, ponlo abajo* (I don't know, put it underneath).
*CHI A: parent…
*CHI B: *¡que es punto ahí abajo!* (I'm telling you it's full stop there underneath!)
*CHI A: *no, ¿por qué? Es el texto, no es palabras* (no, why? It's a text, not [spare] words).
*CHI B: *ya lo sé pero… ¿qué haces? (…) si pones punto, no pongas minúscula.*
(I know but… what are you doing? (…) if you put a full stop don't write small letters)

Next, examples in (22) display discussions related to spelling and punctuation (Mech). In (22a), for instance, the dyad merges comments on these two topics. First, CHI A starts reflecting about the use of a full stop, but CHI B interrupts him to indicate a spelling mistake. Right after CHI A notices his mistake, he formulates the orthography rule applied to double consonants and confirms the use of a full stop at the end of that sentence. In (22b), however, the two children seem to be at odds when discussing the need for full stops, as CHI A considers that the way CHI B has inserted them in the text is unnatural for a text layout. CHI B finishes the discussion upset and he requires his partner to change the words to small letters, since there is no longer a full stop.

(23) Disc

(23a) [ABE030 & ABE042 - Day 1]

*CHI A: *y luego ponemos* (and then we put) "to finally" or to…
*CHI B: to…
*CHI A: *para terminar* (to finish).
*CHI B: to finally, *sí* (yes).
*CHI A: *sí* (yes), to finally…

(23b) [MAR002 & MAR015 - Day 1]

*CHI B: *espera, lo leo* (wait, I'll read it)*:* Eva has sleep and the Eva father has prepared a sandwich, *pero vamos a contar* (but let's count "and" and "Eva") ↑"and" y ↑"Eva" (…) *tres cuatro, cinco*… (three, four, five…) [erases and changes something in the text CHI A helps him to count].

Finally, turning to the last focus category, Disc, example (23a) shows how the two members of the dyad discuss the possibility of including a discourse marker to show the end of the story. However, the one that they decide to use is not totally correct, as the pair decides to translate directly from the L1 equivalent. Conversely, example (23b) shows another use of self-directed talk, as CHI B realizes they have been repeating two items ("Eva" and "and") too much and decides to count the number of instances. Hence, we could describe this episode as an attempt to correct their writing coherence.

Regarding RQ1b, LREs were assessed according to three different criteria: outcome, depth of engagement and use of PKL.

Starting with the first dimension (outcome), results are summarized for all pairs in Table 14 and Figure 5 (the inferential statistics can be found in B1 in APPENDIX B). As can be observed in all four categories, these child learners were able to correctly resolve most LREs. The greatest proportion of target-like resolutions can be found in Mech, and the lowest in Lex, although this percentage was above 46%. The statistical analysis revealed that, on average, learners produced significantly more correct LREs than incorrect, addressed and ignored LREs in F-targets, F-others, Lex and Mech (with effect sizes ranging from small to medium). In the case of Disc, a significant difference was found between correct and addressed LREs (with a small effect size).

Nonetheless, the second most common outcome scenario shows some differences between the focus categories. In F-targets, F-others and Mech this average corresponded to the category of incorrect LREs. Regarding inferential statistics, in F-others LREs, significant differences were found between incorrect and addressed (with a medium effect size), as well as between addressed and ignored (with a small effect size), the average of the latter being higher than the former. In Mech, a significant difference was found between incorrect and addressed, and also between incorrect and ignored (with small effect sizes). However, in the case of Lex, the mean rate of addressed episodes was the second highest, significantly different from incorrect LREs (with a medium effect size). Finally, in Disc, ignored episodes ranked second.

In what follows, we provide examples of each outcome scenario:

(24) Correctly resolved [ABE050 & ABE066 - Day 1]

*CHI A: yes [starts reading], Laura's dad forgot prepare the sandwich to the school, her dad give money to Laura to buy food, he see.
*CHI B: she!
*CHI A: see the black chocolate bar, and buy an… buy the apple she angry and eat everything in the class, her face were red as the teacher call to her father, and then go to the hospital, and for last he…
*CHI B: she!
*CHI A: she see a doctor.

(25) Incorrectly resolved [MAR009 & MAR012 - Day 1]

*CHI B: *vale, aquí llega mi gran pregunta: siempre se me olvida cómo se escribe school, escribo siempre mal ese coool.*
(ok, here comes my biggest question: I always forget how to write 'school', I always write s-cool wrong)
*CHI A: *ese creo que no (,) creo que es ese hache ce* (I don't think it's 's', I think it's s-h-c).
*CHI B: *sí* (yes).
*CHI A: *tiene más sentido* (I think it makes more sense).
*CHI B: *si no sería ↑shool* (otherwise it would be shool*).
*CHI A: *shool jajaja ¿es Laura o Lola?* (shool hahaha, is it Laura or Lola?)

(26) Unresolved addressed [ABE016 & ABE020 - Day 1]

*CHI B: *no, y se puso mala, pon: "y se puso mala".*
(no, and she got sick, write: "and she got sick")
*CHI A: and she… *¿cómo se dice mala?* (how do you say 'sick'?)
*CHI B: [laughs] no sé, a ver piensa en otra cosa.
(I don't know, let's see, think of something else)
*CHI A: *no sé…* (I don't know…)
*CHI B: *bueno pues pon 'alérgica' y ya está* (ok, put 'allergic' and that's it).
*CHI A: *¿y cómo se dice?* (and how do you say it?)
*CHI B: *da igual cómo* (it doesn't matter how).
*CHI A: peanos&(peanuts) *¿y se puso alérgica?* (and she got allergic?)
*CHI B: and, *sí, tía, cómo se dice,* (yes, dude, how do say) it… hit, *no eso es morir* (and, , hit, no, that's to die).
*CHI A: [whispering] the Laura in school in a sandwich put chocolate and peanus&(peanuts).

(27) Unresolved ignored [MAR020 & MAR021 - Day 1]

*CHI B: [crossing out from notes] time… delicious…
*CHI A: *¿cómo se escribe?* (how do you write it?)
*CHI B: *no he puesto nada ni de María, ni de este ni de la madre…*
(I haven't put anything about María or about this or about the mother)
*CHI A: *de los abuelos sí* (about the grandparents you have).

In (24), CHI A is retelling the original story they have written together when CHI B notices a subject pronoun mistake ('he' instead of 'she'). Yet, CHI A does not realize this correction (probably mistaken by the pronunciation of 'see' and 'she') and keeps reading. In the next sentence, CHI A makes the same mistake and CHI B points it out again. This time, CHI B clearly realizes what exactly is wrong and changes the subject pronoun, hence reaching a target-like sentence. In contrast, in (25), CHI B reflects on the spelling of a word (‘school’) that is problematic for her. Her partner provides a solution explaining what he thinks would be a logical pronunciation rule in English, although it does not result in the target-like spelling of the word. CHI A is convinced by this explanation and CHI B finishes the discussion by suggesting what the pronunciation of ‘school’ would be like with an alternative spelling. Hence, the dyad arrives at an incorrect outcome.

The following two examples represent instances in which learners failed to reach any decision on the metalinguistic reflection and, therefore, their collaborative dialogue had no reflection in their written output. For instance, in (26), CHI B wonders how to say ‘get sick’ in English, but as CHI A struggles even with the word ‘sick’, CHI B decides to look for another way of expressing the same idea and suggests the word *‘alérgica’* (allergic). Nevertheless, the pair cannot find the correct equivalent in English and in the end CHI A moves on to another idea from the text. In other words, although both learners actively engage in the lexical search they are unable to find any appropriate solution. In the last example, (27), CHI A, who is writing the story in that moment, asks his peer about the spelling of ‘delicious’, but CHI B concentrates on getting the ideas of the text right that ignores CHI A’s question. Moreover, after CHI B comments on the content, CHI A does not continue challenging the spelling of certain words.

Moving on now to the second classification criterium, depth of engagement, we distinguished between simple and elaborate LREs in each of the topic categories. The descriptive results can be found in Table 15 and the inferential statistics for the mean difference between simple and elaborate in the appendix B1. On average, learners produced more simple than elaborate LREs in all focus categories except for Lex. Moreover, in Lex, the mean difference between elaborate and simple was significant, with a small effect size. Examples of elaborate Lex LREs can be found in (21) and (26), discussed above. In (21b), for instance, both learners engage in giving a definition of ‘lunchbox’, CHI A explains in the L1 what her idea of that word is, and once CHI B gets a hint, he suggests the likeliest equivalent in Spanish, which CHI A approves. In (26) they are both trying to find the L2 equivalent to ‘get sick’, asking questions to each other, although they eventually fail to find the correct expression to convey this idea. In contrast, in (21a), the LRE takes place in a few short turns, as CHI A notices the mistake in CHI B's collocation, and CHI B simply repeats the target-like expression, without asking further or showing any attempt to delve into this linguistic matter.

To finish the analysis of LREs on Day 1, we will now summarize the results regarding the use of PKL and L2 in learners' collaborative dialogue involving metalinguistic discussions. Table 16 summarizes the descriptive statistics and appendix B1 contains the inferential paired-samples t-tests. As can be observed in the mean results and raw proportions for the number of turns, the use of either PKL or L2 varied according to the topic of discussion. On average, learners generated more turns in the L2 than in PKL in order to reflect on grammar (F-targets and, especially, F-others), whilst the opposite was true regarding Lex, Mech and Disc. The difference reached significance in the case of F-others and Mech (with small effect sizes). The results from the comparison of the number of words in PKL and L2 reflected the same trend, although the only significant differences were found in Mech and Disc. The only exception to the similarity of PKL use in turns and words was found in F-targets, where the PKL word average was higher than the L2 word average.

In order to have a better estimation of the length of the turns in PKL and L2, we divided the number of words by the number of turns in each of the language categories (Table 17). On average, learners produced longer turns in PKL than in the L2 in F-targets, Lex and Disc, and the mean difference proved to be statistically significant in the case of F-targets and Disc (with small effect sizes). Conversely, the mean length of L2 turns was longer than PKL turns in the case of F-others and Mech, but it failed to reach statistical significance. Taking all focus categories together, the average length of PKL and L2 turns was virtually identical. The following two examples from the same pair show how their use of PKL and L2 changes depending on the linguistic topic they were discussing:

(28) [ABE012 & ABE022 - Day 1]

(28a) PKL in Mech

*CHI B: yes, Laura…
*CHI A: Laura?
*CHI B: *ah, pues yo no sé cómo se llama, yo he entendido Laura.*
(oh, I don't know what she's called, I understood Laura)
*CHI A: is Lora! [total sound spelling correspondence]
*CHI B: *es que Lora en inglés se dice, se escribe Laura.*
(that's because Lora in English is written Laura)
*CHI A: ah, ok, ok [giggles].

(28b) L2 in F-others

*CHI B: and chocolate.
*CHI A: but… eh…. I shop the chocolate.
*CHI B: and they shop the apple and chocolate.
*CHI A: chocolate and eh… *cacahuet*&(peanut).
*CHI B: *¿cacahuete?* (peanut) [desperate] umm (…) but Laura, but Laura…
*CHI A: no Laura.
*CHI B: is.
*CHI A: no Laura I go to the class.
*CHI B: but Laura go to class.
*CHI A of Maths.

In this dyad, CHI B seems to have a stronger command of the L2 than his peer. However, his way of providing feedback varies. In (28a), CHI A is convinced that the main character's name is 'Lora' (influenced by the fact that Spanish has a transparent orthography, with a total letter-sound correspondence). However, CHI B explains in Spanish that the spelling of that word is equivalent to the L1 female name of 'Laura'. However, in (28b), CHI B notices that CHI A is inserting a superfluous subject pronoun 'I' between the subject and the verb. The strategy this time is to recast his peer's sentence by getting rid of the extra first-person singular pronoun. Unlike in the previous example, where CHI A is completely aware of his mistake, in (28b) the feedback seems to go unnoticed by CHI A, as he simply completes the previous sentence ("of Maths").

Table 14 Outcome of LREs / focus

	Resolved								Unresolved							
	Correctly				Incorrectly				Addressed				Ignored			
	N	M	SD	95% CI	N	M	SD	95% CI	N	M	SD	95% CI	N	M	SD	95% CI
F-targets	19	0.65	1.23	[0.19-1.12]	7	0.24	0.63	[-0.01-0.48]	2	0.07	0.26	[-0.03-1.17]	5	0.17	0.47	[-0.01-0.35]
F-others	67	2.31	2.12	[1.5-3.11]	34	1.17	1.28	[0.68-1.66]	9	0.31	0.66	[0.06-0.56]	26	0.89	0.9	[0.55-1.24]
Lex	64	2.21	1.72	[1.55-2.86]	14	0.48	0.95	[0.12-0.84]	35	1.21	1.42	[0.66-1.74]	22	0.76	1.06	[0.36-1.16]
Mech	132	4.55	2.64	[3.55-5.56]	34	1.17	1.1	[0.75-1.59]	10	0.34	0.67	[0.09-0.59]	12	0.41	0.63	[0.17-0.65]
Disc	13	0.45	0.83	[0.13-0.76]	5	0.17	0.47	[-0.01-0.35]	2	0.07	0.26	[-0.03-0.17]	7	0.24	0.69	[-0.02-0.5]
Total	295	10.17	6.3	[7.77-12.57]	94	3.24	2.15	[2.42-4.06]	58	2	2.2	[1.16-2.83]	72	2.48	2.54	[1.51-3.45]

Figure 5 Outcome percentages / focus

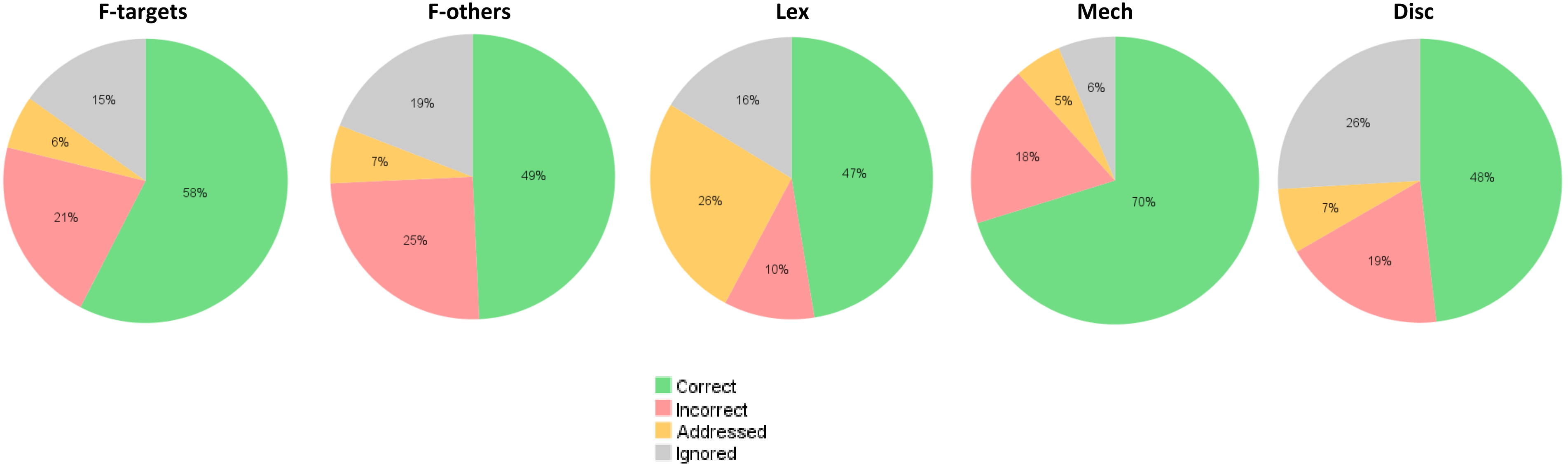


Table 15 LRE engagement / focus

	Simple				Elaborate			
	N	**M**	**SD**	**95% CI**	**N**	**M**	**SD**	**95% CI**
F-targets	19	0.65	1.45	[0.11-1.21]	14	0.48	0.95	[0.12-0.84]
F-others	79	2.72	1.77	[2.01-3.40]	57	1.96	2.14	[1.15-2.78]
Lex	43	1.48	1.45	[0.93-2.04]	92	3.17	2.20	[2.33-4.01]
Mech	100	3.45	2.13	[2.64-4.26]	88	3.03	2.54	[2.07-4]
Disc	15	0.52	1.09	[0.10-0.93]	12	0.41	0.57	[0.20-0.63]
Total	256	8.83	5.34	[6.80-10.86]	263	9.07	5.34	[6.99-11.14]

Figure 6 LRE engagement percentages / focus

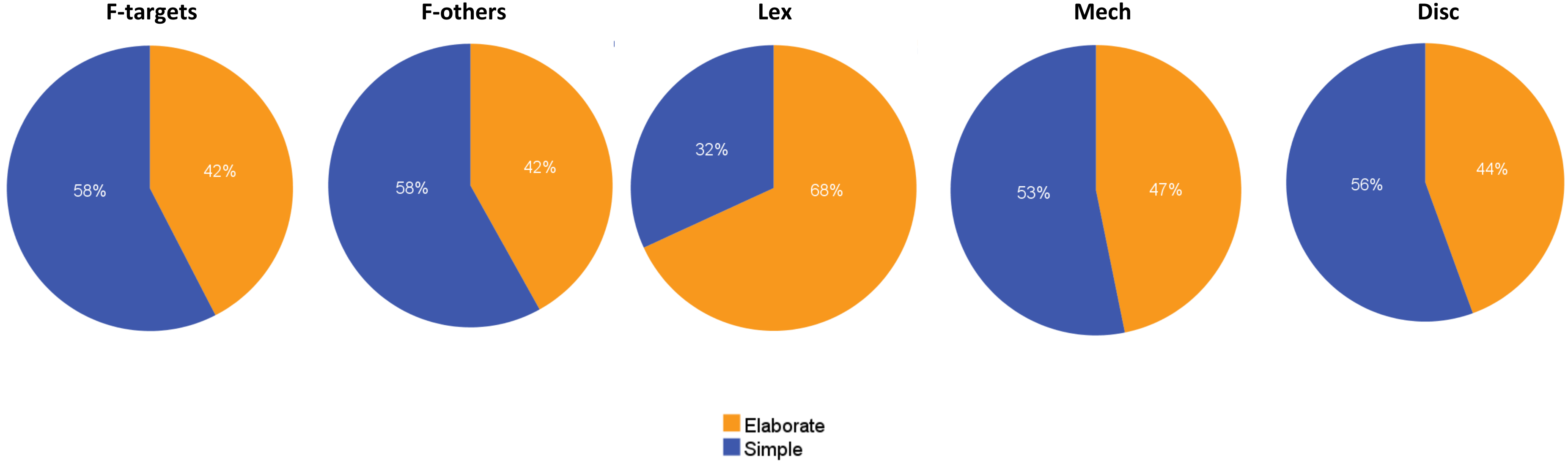


Table 16 LRE turns and length (in words) in PKL and L2 / focus

	PKL								L2							
	Turns				Words				Turns				Words			
	N	M	SD	95% CI	N	M	SD	95% CI	N	M	SD	95% CI	N	M	SD	95% CI
F-targets	58	2	5.91	[-0.25-4.25]	526	18.14	52.42	[-1.80-38.08]	79	2.72	3.42	[1.42-4.02]	473	16.31	33.14	[3.70-28.92]
F-others	116	4	6.54	[1.51-6.49]	1003	34.59	60.35	[11.63-57.54]	269	9.27	8.40	[6.08-12.47]	1648	56.83	56.18	[35.45-78.2]
Lex	277	9.55	8.56	[6.29-12.81]	1655	57.07	48.9	[38.47-75.67]	220	7.59	6.46	[5.13-10.04]	1134	39.10	41.57	[23.29-54.92]
Mech	302	10.41	8.72	[7.10-13.73]	1948	67.17	60.71	[44.08-90.27]]	135	4.66	3.50	[3.32-5.99]	1026	35.38	41.13	[19.73-51.03]
Disc	46	1.59	2.65	[0.58-2.60]	411	14.17	25.05	[4.64-23.70]	19	0.66	1.11	[0.23-1.08]	73	2.52	4.81	[0.69-3.35]
Total	799	27.55	18.60	[20.47-34.63]	5543	191.14	147.28	[135.12-247.16]	722	24.90	14.59	[19.34-30.45]	4354	150.14	121.62	[103.85-196.4]

Figure 7 Proportion of PKL turns vs L2 turns

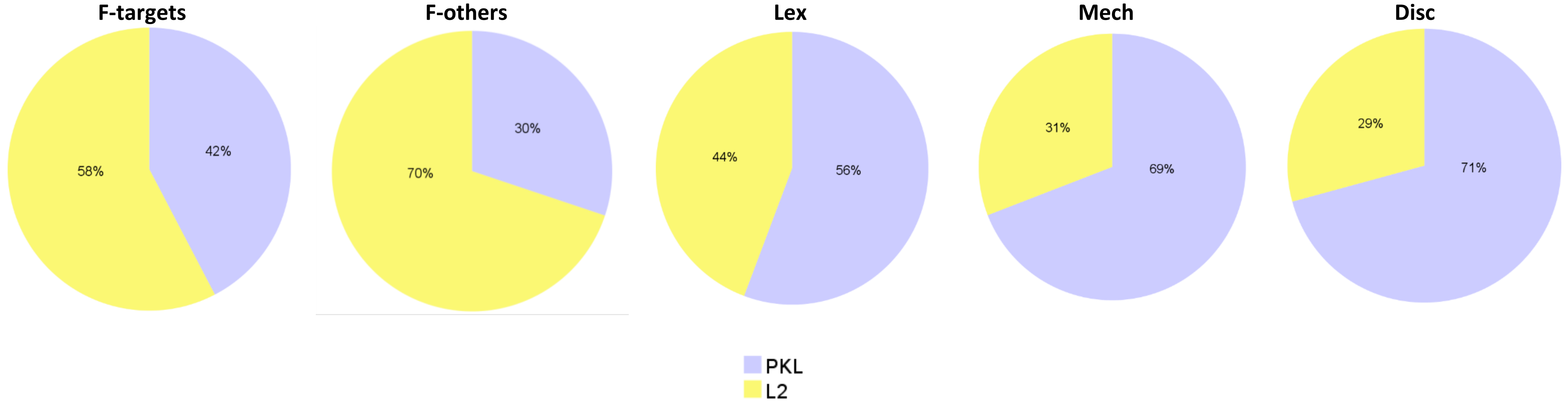


Table 17 Turn length in PKL and L2 (words/turns) / focus and paired-samples t-tests

	M	SD	95% CI	PKL vs L2			
				t	df	*p*	*d*
F-Target PKL length	3.04	5.25	[1.66-4.13]	3.21	57	0.002	0.42
F-Target L2 length	1.41	4.30	[0.09-1.82]				
F-others PKL length	5.90	6.66	[3.83-6.90]	-1.92	57	0.059	-0.25
F-others L2 length	6.81	4.36	[6.09-8.69]				
Lex PKL length	5.62	3.26	[4.16-6.03]	0.94	57	0.35	0.12
Lex L2 length	4.83	2.86	[3.77-5.28]				
Mech PKL length	5.83	4.47	[3.88-5.89]	-1.86	57	0.068	-0.24
Mech L2 length	6.97	5.01	[4.92-7.50]				
Disc PKL length	4.12	6.78	[2.54-5.43]	2.88	57	0.005	0.38
Disc L2 length	1.46	2.51	[0.96-2.63]				
TOTAL L1 length	6.15	2.77	[5.18-6.73]	-0.61	57	0.547	-0.08
TOTAL L2 length	6.18	3.63	[5.45-7.04]				

RQ1c is related to the analysis of the impact of task-related variables, more specifically, TR and pretask FFI. First, the results from the written languaging (wLREs) on Day 2 are reported for the Comp group in Table 18. The sample of participants on this occasion was also reduced as four child learners (different from the missing participants on Day 1) failed to carry out the dictogloss task on Day 2:

Table 18 wLREs on Day 2/ focus

	Comp (n = 30)			
	N (%)	**M**	**SD**	**95% CI**
F-targets	0	0	0	-
F-others	1 (4.3)	0.03	0.18	[-0.03 - 0.10]
Lex	8 (34.8)	0.27	0.52	[0.07 - 0.46]
Mech	13 (56.5)	0.43	0.70	[0.16 - 0.70]
Disc	1 (4.3)	0.03	0.18	[-0.03 - 0.10]
Total	23	0.77	1.16	[0.33 - 1.20]

As can be observed, on Day 2, learners who carried out the DG task individually primarily referred in their posttask metalinguistic reflections to Mech and Lex. However, the number of wLREs decreased as compared to Day 1 (although it failed to reach statistical significance, as can be seen B2, in APPENDIX B). Furthermore, target forms were still not discussed. In general, learners expressed their comments in similar terms to those of Day 1: *"se me ha olvidado cómo se dice 'piscina'"* (I forgot how to say 'swimming pool') [ABE062 - Day 2]; *"¿cómo se dice 'play' en pasado?"* (how do you say 'play' in the past?) [ABE029 - Day 2]; *"he escrito 'haiscrim' así pero no estoy segura de que esté bien"* (I have written 'haiscrim', but I am not sure whether it's fine); *"'hour' idatzi dut 'ordua' bezala eta ez dakit horrela idazten den"* (I wrote 'hour' to express *'ordua'* but I don't know if it's correct) [ABE051 - Day 2]. Finally, as was the case on Day 1, child learners on Day 2 devoted many of their written comments to express their views on non-linguistic issues, such as their perceived task difficulty or their doubts about the content: *"no sé qué es lo que celebran"* (I don't know what they are celebrating) [ABE005 - Day 2]; *"no sé si en el texto se decía que hacía mucho calor"* [ABE009 - Day 2]; *"escribir esta historia ha sido más fácil"* (it has been easier to write this story) [ABE034 - Day 2]; *"me ha costado entender a la chica que hablaba y he escrito* *algo"* (I struggled to understand the girl who was narrating the story and I wrote something) [ABE004 - Day 2].

We will now turn to compare the results of languaging generated as a result of the collaborative DG on Day 1 and Day 2. A series of repeated-measures ANOVA with Bonferroni adjustment for multiple comparisons were run on JASP (JASP Team, 2019) in order to ascertain the impact of TR, the instruction type (Collab vs FFI+Collab), as well as the potential interaction between TR (Time) ⁎ Instruction Type. In order to interpret the eta square (η^2) effect sizes derived from the analyses, we were guided by Plonsky and Oswald (2014, p. 894), who recommended a procedure for calculating the typical L2 research benchmark values out of the *r* value percentiles found in their meta-analytic study. Hence, the eta square thresholds were the following: small: $\eta^2 = .06$; medium: $\eta^2 = .16$; large: $\eta^2 = .36$.

Firstly, the time spent on the task on Day 1 and Day 2 by instruction type condition is reported (Table 19 and Figure 8). The statistical analysis showed that the interaction Time ⁎ Instruction was significant, yet with a small effect size ($F = 5.68$, $p = .02$, $\eta^2 = .05$). On Day 1, FFI+Collab spent significantly more time on the task than Collab (t = 5.96, $p < .001$, CI [294.2 – 796.16]), and that was also the case on Day 2 (t = 2.88, $p = .03$, CI [12.66, 514.62]). Moreover, with regards to intragroup differences, FFI+Collab spent significantly less time on Day 2 than on Day 1 (t = -5.78, $p = .016$, CI [-37.77, -505.17]), while the analysis failed to show any significant change in time-on-task in Collab.

Regarding the impact of TR and instruction type on LRE focus, Table 20 and Table 22 contain the descriptive statistics for LREs and turns on Day 1 in each treatment condition, and Table 21 and Table 23 contain the results with regards to Day 2. Starting with the intergroup comparison, on average, on Day 1 learners in FFI+Collab generated more LREs than Collab in all focus categories except for F-others, and regarding turns, the former also generated more in all focus categories. Nevertheless, on Day 2, the average of LREs showed some differences, as Collab produced more than FFI+Collab in F-others, Mech, and Disc (and viceversa in the case of F-targets and Lex). Yet, the average number of turns on Day 2 was still larger in FFI+Collab in all topic categories except for Disc.

Examining the intragroup differences, it can be observed that the mean LRE rates varied slightly from Day 1 to Day 2 in each of the groups. Collab increased the LRE average in F-others and D, whereas FFI+Collab did so in F-targets and F-others. The rest of LRE rates remained very similar between Day 1 and Day 2, with the exception of Mech in FFI+Collab, where the mean decreased by more than 2 points. The average number of turns also reflected stability between the two days within the two groups. Learners in Collab produced more Disc turns on Day 2, whilst learners in FFI+Collab produced more Lex turns. The sharpest decrease corresponded again to the category of Mech in FFI+Collab (an average of 18 turns on Day 1 vs 12 on Day 2).

A series of repeated-measures ANOVA were run for each of the topic categories, and the interaction plots resulting from those analyses can be found in Figure 9 (for the complete inferential statistics, see B2). The analyses failed to show any significant effects of time, instruction type or interactions. In other words, dyads were similar in their attention to form regardless of the day and their instruction type. However, the exception to this was found in the case of Lex, where the analysis indicated a significant main effect of instruction type with a medium effect size (F = 5.39, p = .028, η^2 = .13). The post-hoc analyses revealed that on average FFI+Collab discussed lexis significantly more than Collab, regardless of testing time (t = -2.32, p = .03, CI [0.26, 4.22]), however the CI indicate that the mean difference was close to the zero value, which implies that even in this case the difference between the two groups was unimportant.

As far as LRE outcome is concerned, the descriptive statistics for each condition per treatment day can be found in Table 24 (Day 1) and Table 25 (Day 2). The mean rates remained very similar on both days. In the category of “Correctly”, however, FFI+Collab learners on Day 1 produced more than their counterparts in Collab, but the difference was inverted on Day 2. With regards to “Incorrectly”, Collab decreased the mean rate from Day 1 to Day 2, whereas FFI+Collab produced virtually the same average. Yet, the repeated-measures analyses failed to show any significant differences (see Figure 10 displaying the interaction plots).

Turning to the analysis of LRE engagement, the descriptive statistics for each condition can be found in Table 26 (Day 1) and Table 27 (Day 2). Regarding intragroup differences, it can be observed that both FFI+Collab and Collab decreased the mean number of elaborate LREs from Day 1 to Day 2, while the opposite was true for simple LREs. Examining between-group comparisons, Collab dyads produced more simple LREs than FFI+Collab on both days, and the latter generated more elaborate LREs than the former on both occasions. The statistical analysis (see B2 and Figure 11 for the interaction plots) revealed a significant main effect of the Instruction type for elaborate LREs, with a medium effect size (F = 5.19, p = .03, η^2 = .13). The post-hoc analysis showed that FFI+Collab, regardless of the treatment day, produced a significantly higher average of elaborate LREs than Collab (t = 2.28, p = .03, CI [0.39, 7.52]). Conversely, we failed to find any significant effects or interaction in the case of simple LREs.

As far as the use of PKL and L2 is concerned, the mean length of the turns in each language was analyzed with regards to time and instruction type (Table 28 and Table 29 display the descriptive statistics). The intragroup comparison shows that Collab produced longer L2 than PKL turns on both days, while the opposite was true for FFI+Collab. From Day 1 to Day 2, Collab increased the average length of L2 turns, whereas that of PKL turns decreased (and vice versa in the case of FFI+Collab). Examining between-group comparisons, on both days Collab yielded on average longer L2 turns than FFI+Collab, but FFI+Collab produced longer PKL turns. The statistical analyses (see B2 and Figure 12 for the interaction plots) revealed a significant main effect of instruction type in PKL turns with a medium effect size (F = 8.23, p = .008, η^2 = .15). The post-hoc analyses showed that, regardless of the day, FFI+Collab produced on average significantly longer turns in PKL than Collab (t = 2.87, p= .008, CI [0.64, 3.84]). Conversely, the same analysis with regards to L2 failed to show any significant difference.

Table 19 Time on task and impact of TR / experimental group

	Day 1					Day 2				
	M	**SD**	**95% CI**	**Min**	**Max**	**M**	**SD**	**95% CI**	**Min**	**Max**
Collab (n = 14)	00:10:37	00:04:47	[00:07:51-00:13:23]	00:03:20	00:18:36	00:10:47	00:03:44	[00:08:37-00:12:57]	00:04:28	00:17:47
FFI+Collab (n = 15)	00:19:42	00:01:11	[00:17:08-00:22:16]	00:12:56	00:27:58	00:15:06	00:03:03	[00:13:29-00:16:52]	00:09:59	00:21:39

Figure 8 Interaction plot showing time-on-task (in seconds) for treatment day and experimental group (error bars represent 95% CI)

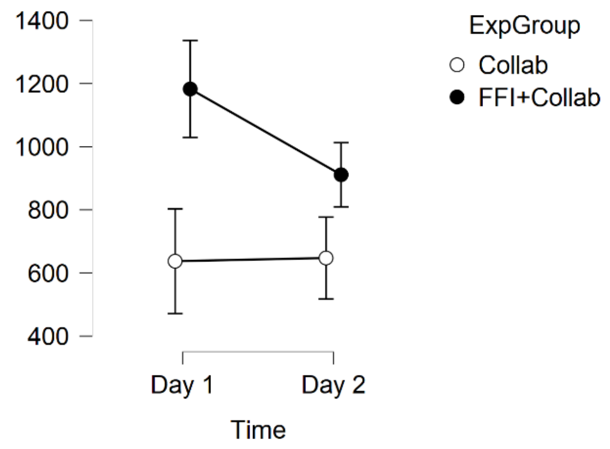


Table 20 LREs on Day 1 / experimental group

	Collab (n = 14)				FFI+Collab (n = 15)			
	N	**M**	**SD**	**95% CI**	**N**	**M**	**SD**	**95% CI**
F-targets	12	0.86	1.66	[-0.09-1.81]	21	1.40	2.41	[0.06-2.73]
F-others	67	4.78	3.64	[2.68-6.89]	69	4.60	3.06	[2.90-6.29]
Lex	49	3.50	2.41	[2.11-4.89]	86	5.73	2.96	[4.09-7.37]
Mech	79	5.64	2.68	[4.09-7.19]	109	7.27	3.89	[5.11-9.43]
Disc	8	0.57	0.85	[0.08-1.06]	19	1.27	1.75	[0.29-2.24]
Total	215	15.36	8.59	[10.40-20.31]	304	20.27	9.08	[15.23-25.29]

Table 21 LREs on Day 2 / experimental group

	Collab (n = 14)				FFI+Collab (n = 15)			
	N	**M**	**SD**	**95% CI**	**N**	**M**	**SD**	**95% CI**
F-targets	10	0.71	0.99	[0.14-1.29]	33	2.20	2.14	[1.01-3.39]
F-others	81	5.78	4	[3.47-8.1]	72	4.80	3.28	[2.98-6.61]
Lex	45	3.21	2.04	[2.03-4.4]	82	5.47	4.17	[3.16-7.78]
Mech	79	5.64	4.12	[3.26-8.02]	76	5.07	3.99	[2.86-7.28]
Disc	17	1.21	1.48	[0.36-2.07]	15	1	1	[0.45-1.55]
Total	232	16.57	9.20	[11.26-21.89]	278	18.53	10.74	[12.59-24.48]

Table 22 Turns on Day 1 / experimental group

	Collab (n = 14)				FFI+Collab (n = 15)			
	N (%)	M	SD	95% CI	N (%)	M	SD	95% CI
F-targets	46	3.28	3.91	[1.03-5.54]	91	6.07	8.89	[1.14-10.99]
F-others	163	11.64	11.23	[5.16-18.13]	222	14.80	14.39	[6.83-22.77]
Lex	179	12.78	10.19	[6.89-18.67]	318	21.20	13.94	[13.48-28.92]
Mech	163	11.64	6.96	[7.63-15.66]	274	18.27	12.77	[11.19-25.34]
Disc	19	1.36	1.82	[0.30-2.41]	46	3.07	4.01	[0.85-5.29]
Total LRE-turns	570	40.71	21.15	[28.50-52.92]	951	63.40	26.75	[48.58-78.21]
Total turns	1237	88.36	35.06	[67.86-108.86]	2316	154.40	46.94	[128.41-180.39]

Note: Total turns (with and without LREs)

Table 23 Turns on Day 2 / experimental group

	Collab (n = 14)				FFI+Collab (n = 15)			
	N	M	SD	95% CI	N	M	SD	95% CI
F-targets	33	2.38	4.62	[-0.31-5.02]	80	5.33	5.46	[2.31-8.36]
F-others	147	10.5	8.3	[5.71-15.29]	171	11.4	8.32	[6.79-16.01]
Lex	138	9.86	8.79	[4.78-14.93]	354	23.6	18.12	[13.56-33.63]
Mech	161	11.5	9.15	[6.21-16.78]	179	11.93	10.66	[6.03-17.84]
Disc	48	3.43	4.09	[1.07-5.79]	44	2.93	3.79	[0.83-5.03]
Total LRE-turns	527	37.64	21.6	[25.17-50.11]	828	55.2	31.49	[37.76-72.64]
Total turns	1209	86.36	44.13	[60.87-111.84]	1503	100.2	42.56	[76.63-123.77]

Note: Total turns (with and without LREs)

Figure 9 Interaction plots showing LRE focus by day and experimental group (error bars represent 95% CI)

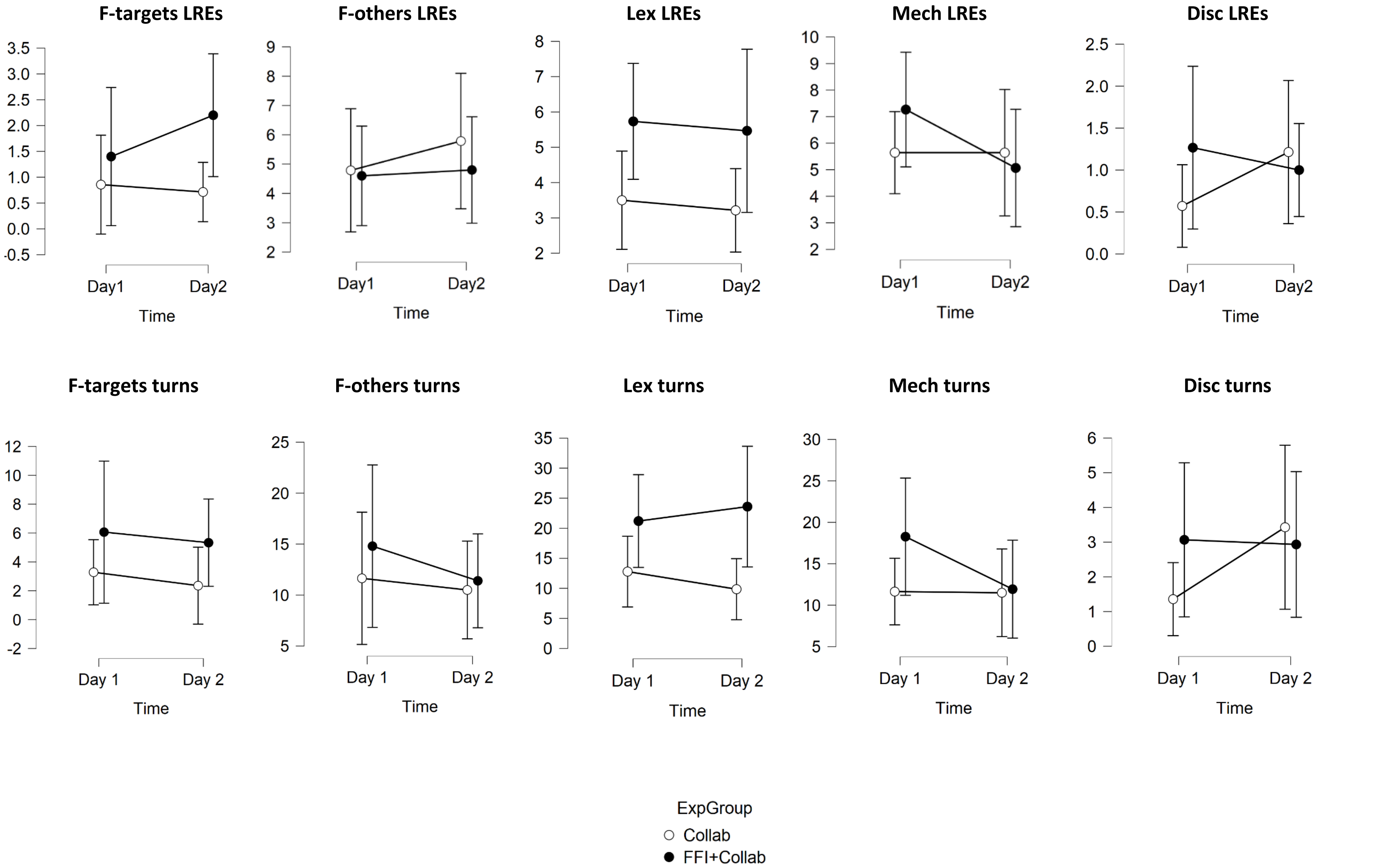


Table 24 Outcome of LREs on Day 1 / experimental group

	Correct		Incorrect		Addressed		Ignored	
	Collab (n = 14)	FFI+Collab (n = 15)	Collab (n = 14)	FFI+Collab (n = 15)	Collab (n = 14)	FFI+Collab (n = 15)	Collab (n = 14)	FFI+Collab (n = 15)
N	121	174	43	51	19	39	32	40
M	8.64	11.60	3.07	3.40	1.36	2.60	2.29	2.67
SD	4.53	7.47	1.68	2.56	1.39	2.67	3.27	1.72
95% CI	[6.02-11.26]	[7.46-15.74]	[2.09-4.04]	[1.98-4.81]	[0.55-2.16]	[1.12-4.08]	[0.4-4.17]	[1.71-3.62]

Table 25 Outcome of LREs on Day 2 / experimental group

	Correct		Incorrect		Addressed		Ignored	
	Collab (n = 14)	FFI+Collab (n = 15)	Collab (n = 14)	FFI+Collab (n = 15)	Collab (n = 14)	FFI+Collab (n = 15)	Collab (n = 14)	FFI+Collab (n = 15)
N	157	161	25	54	13	26	37	37
M	11.21	10.73	1.78	3.60	0.93	1.73	2.64	2.47
SD	6.25	8.16	1.53	2.50	1	1.83	2.31	1.60
95% CI	[7.6-14.82]	[6.21-15.25]	[0.9-2.67]	[2.21-4.98]	[0.35-1.5]	[0.72-2.75]	[1.31-3.97]	[1.58-3.35]

Figure 10 Interaction plots showing LRE outcome by day and experimental group (error bars represent 95% CI)

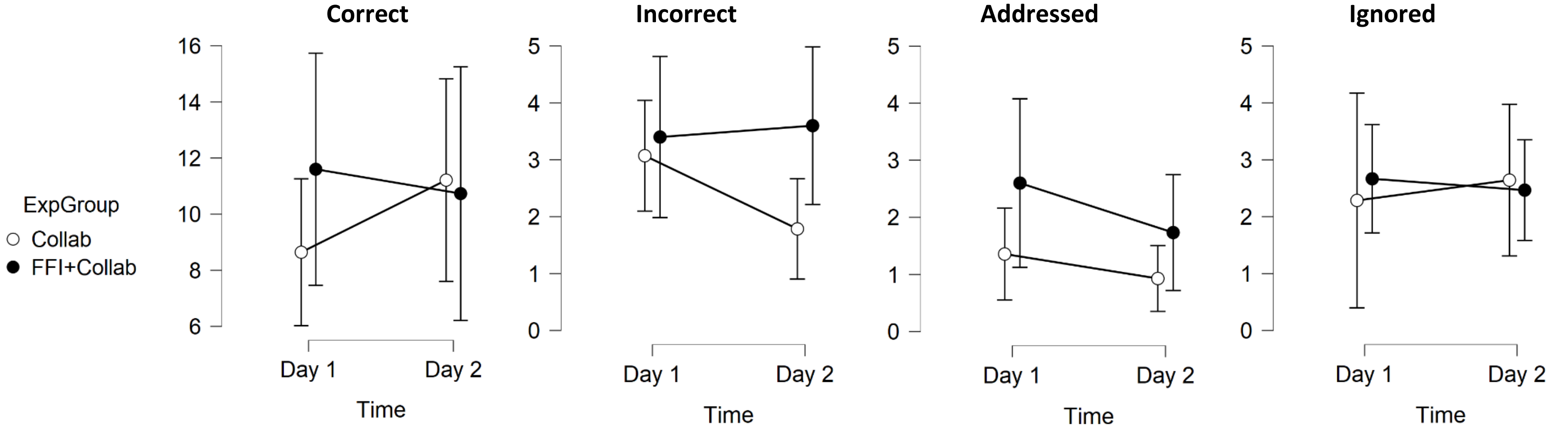


Table 26 Engagement of LREs on Day 1 / experimental group

	Simple		Elaborate	
	Collab (n = 14)	**FFI+Collab (n = 15)**	**Collab (n = 14)**	**FFI+Collab (n = 15)**
N	127	129	88	175
M	9.07	8.60	6.28	11.67
SD	6.08	4.75	4.34	5.20
95% CI	[5.56-12.83]	[5.97-11.23]	[3.78-8.79]	[8.78-14.55]

Table 27 Engagement of LREs on Day 2 / experimental group

	Simple		Elaborate	
	Collab (n = 14)	**FFI+Collab (n = 15)**	**Collab (n = 14)**	**FFI+Collab (n = 15)**
N	147	149	85	129
M	10.50	9.93	6.07	8.60
SD	6.34	5.31	4.18	6.60
95% CI	[6.84-14.16]	[6.99-12.87]	[3.66-8.48]	[4.94-12.25]

Figure 11 Interaction plots showing LRE engagement by day and experimental group (error bars represent 95% CI)

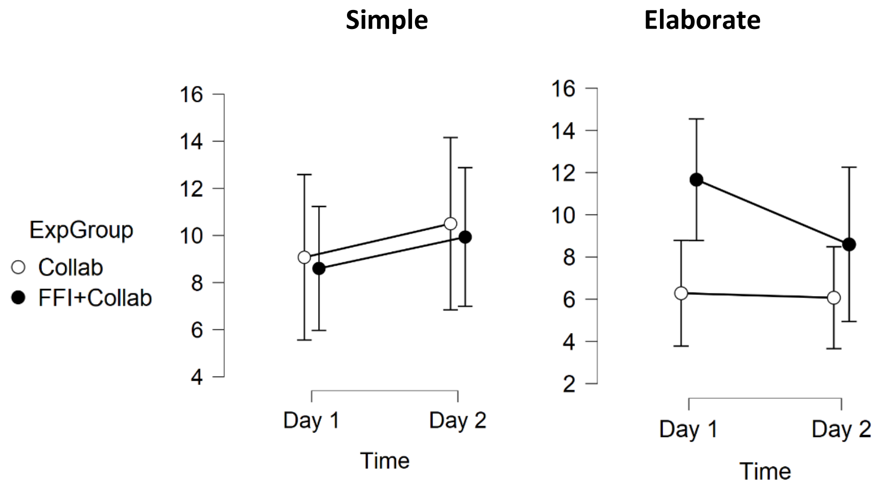


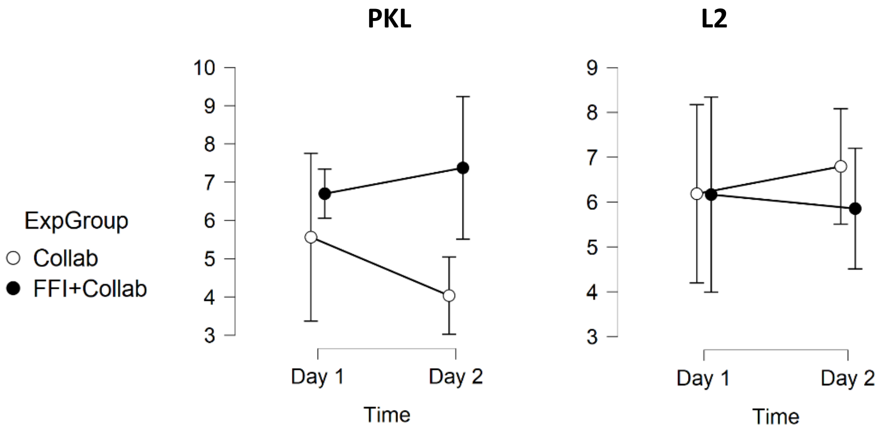
Table 28 Mean length of PKL and L2 turns on Day 1/ experimental group

	PKL turn mean length		L2 turn mean length	
	Collab (n = 14)	FFI+Collab (n = 15)	Collab (n = 14)	FFI+Collab (n = 15)
M	5.56	6.73	6.19	6.17
SD	3.79	1.16	3.44	3.92
95% CI	[3.37-7.75]	[6.06-7.34]	[4.2-8.17]	[3.99-8.34]

Table 29 Mean length of PKL and L2 turns on Day 2/ experimental group

	PKL turn mean length		L2 turn mean length	
	Collab (n = 14)	FFI+Collab (n = 15)	Collab (n = 14)	FFI+Collab (n = 15)
M	4.03	7.37	6.79	5.86
SD	1.75	3.37	2.23	2.43
95% CI	[3.02-5.04]	[5.51-9.24]	[5.51-8.08]	[4.51-7.20]

Figure 12 Interaction plots showing the mean length of PKL and L2 turns by day and experimental group (error bars represent 95% CI)



Finally, we analyzed to what extent learners had focused their attention on each of the target forms (3S and POSS), and more specifically, whether the specially designed DG texts (DG1 and DG2) had succeeded in drawing YLs' attention to these target forms. In order to obtain these results, we first distinguished the LREs related to each of the target forms depending on the original dictogloss text dyads had received (whose order had been counterbalanced on Day 1 and Day 2). That is, if a dyad was working on DG1 and had focused on the 3S, that LRE was counted as a 3S target instance. Conversely, if while working with the same text, they had focused on POSS, that LRE was counted as a POSS non-target instance (and vice versa in the case of DG2). The analysis of the results for this research question was conducted regardless of time (i.e. LREs from Day 1 and Day 2 were counted together), but taking into account the dyad's instruction type (Collab vs FFI + Collab), as this would allow us to see if the pretask FFI stage had boosted learners' attention towards the target forms in DG1 and DG2.

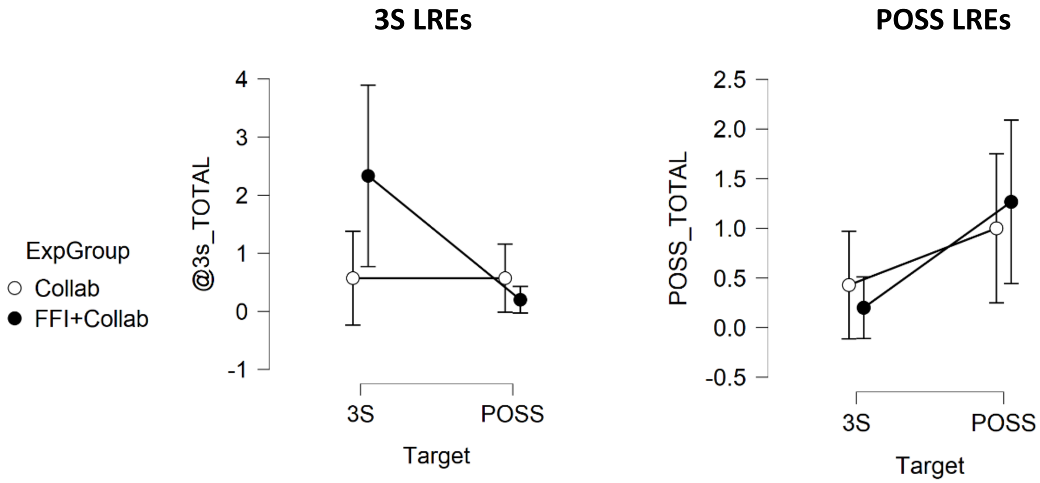
Table 30 summarizes the results, and Figure 13 displays the interaction plots resulting from the two-way ANOVA (see B3 in APPENDIX B for the complete analysis). Looking at the main effects, both in the case of 3S and POSS the analyses revealed a significant effect of the target form (3S: F = 5.83, p = .019, η^2 = 0.09 / POSS: F = 7.63, p = .008, η^2 = 0.12), with effect sizes ranging from small to medium. In other words, irrespective of the treatment condition, learners significantly produced more LREs about 3S and POSS when these were targeted by the DG texts than when they were not. If we introduce the variable of instruction type, some differences can be observed in the performance of FFI+Collab and Collab. With regards to 3S, when it was discussed as a target form (i.e. as part of DG1), FFI+Collab produced significantly more episodes than Collab (t = 2.82, p = .04, CI [0.11, 3.42]); on the contrary, when it was discussed as a non-target form (i.e. while completing DG2), the analysis failed to find any significant differences, that is, FFI+Collab and Collab performed similarly in that case. Turning to POSS, when it was discussed as a target form (i.e. during DG2), FFI+Collab produced slightly more episodes than Collab, but it was far from reaching statistical significance. Conversely, when it was not a target-form (i.e. as part of DG1), the average LRE production was very low in both instruction conditions, but Collab generated on average slightly more episodes (again, not reaching statistical significance).

To conclude, we also checked for within-group differences regarding attention to the 3S and POSS by means of two paired-samples t-tests. However, the statistical analysis failed to show any significant differences and the effect sizes were small ($d < 0.40$).

Table 30 LREs related to 3S and POSS, when target and non-target / experimental group

	3S				**POSS**			
	DG Target		**DG Non-target**		**DG Target**		**DG Non-target**	
	Collab (n = 14)	**FFI+Collab (n = 15)**	**Collab (n = 14)**	**FFI+Collab (n = 15)**	**Collab (n = 14)**	**FFI+Collab (n = 15)**	**Collab (n = 14)**	**FFI+Collab (n =15)**
N	8	35	8	3	14	19	6	3
M	0.57	2.33	0.57	0.20	1	1.27	0.43	0.20
SD	1.39	2.82	1.02	0.41	1.30	1.49	0.94	0.56
95% CI	[-0.33 - 1.47]	[1.46 - 3.2]	[-0.33 - 1.47]	[-0.67 - 1.07]	[0.39 - 0.78]	[0.68 - 1.85]	[-0.18 - 1.03]	[-0.38 - 0.78]

Figure 13 Interaction plots showing 3rd -s and POSS LREs with regards to dictogloss target form and experimental group



6.2 RQ2 Collaborative work

In what follows, the results in response to the second set of RQs are provided. These questions deal with the nature of collaborative work and its impact on the quality of the DG text reconstruction and subsequent individual L2 writing.

a) What are YLs' patterns of interaction during collaborative dictogloss? Are they influenced by TR?
b) To what extent do collaboration, TR and FFI influence the quality of the dictogloss written product?
c) Does dictogloss have an impact on subsequent individual writing?

In order to determine the nature of the dyadic behavior during children's DG performance, their interaction on Day 1 and Day 2 was holistically analyzed and the most predominant pattern of interaction was identified for each dyad (Butler & Zeng, 2015; Storch, 2002, 2009a, 2016). The pattern classification corresponding to Day 1 and Day 2 for each dyad from the two collaborative settings is displayed in

Table 31 (Collab) and Table 32 (FFI+Collab). In addition, Figure 14 and Figure 15 illustrate the proportion of patterns on Day 1 and Day 2 within each experimental condition.

Table 31 Patterns of interaction in Collab on Day 1 and 2

Dyad	Day 1 pattern	Day 2 pattern
ABE003 & ABE010	Collaborative	Collaborative
ABE007 & ABE011	Collaborative	Collaborative
ABE008 & ABE013	Collaborative	Cooperative or passive/parallel
ABE012 & ABE022	Collaborative	Collaborative
ABE016 & ABE020	Collaborative	Dominant/dominant
ABE030 & ABE042	Collaborative	Collaborative
ABE031 & ABE039	Collaborative	Collaborative
ABE032 & ABE036	Cooperative or passive/parallel	Cooperative or passive/parallel
ABE033 & ABE037	Collaborative	Collaborative
ABE050 & ABE066	Collaborative	Collaborative
ABE052 & ABE064	Cooperative or passive/parallel	Collaborative
ABE053 & ABE056	Cooperative or passive/parallel	Cooperative or passive/parallel
ABE054 & ABE061	Collaborative	Collaborative
ABE058 & ABE059	Dominant/passive	Collaborative

Note: Shading is used to indicate the changes in pair dynamics from Day 1 to Day 2.

Figure 14 Proportions of patterns of interaction in Collab on Day 1 and Day 2

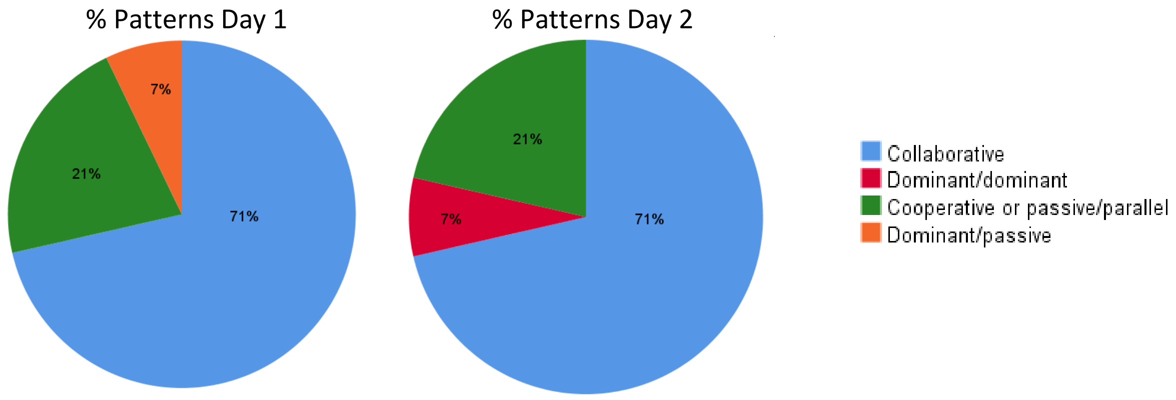
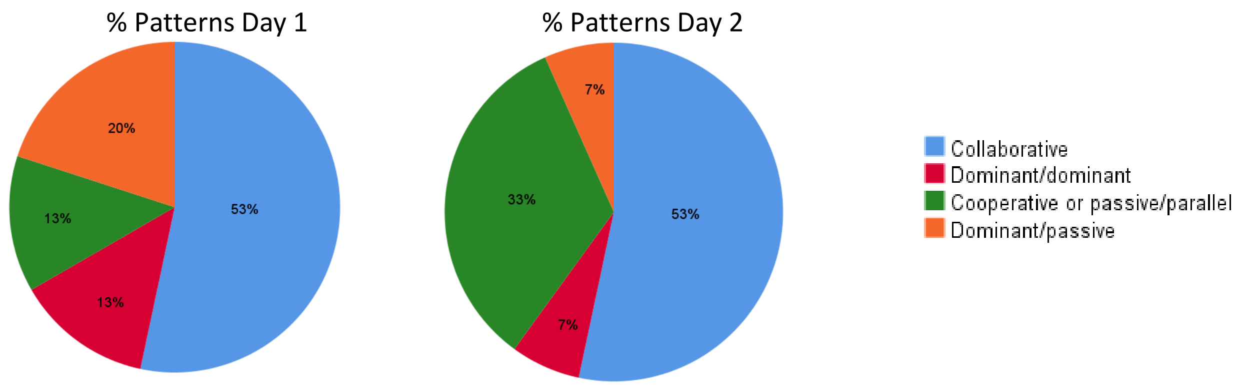


Table 32 Patterns of interaction in FFI+Collab on Day 1 and 2

Dyad	Day 1 pattern	Day 2 pattern
MAR001 & MAR014	Dominant/dominant	Dominant/dominant
MAR002 & MAR015	Dominant/passive	Cooperative or passive/parallel
MAR003 & MAR007	Collaborative	Collaborative
MAR004 & MAR019	Collaborative	Collaborative
MAR006 & MAR018	Collaborative	Collaborative
MAR008 & MAR017	Collaborative	Collaborative
MAR009 & MAR012	Collaborative	Collaborative
MAR011 & MAR016	Dominant/dominant	Dominant/passive
MAR020 & MAR021	Dominant/passive	Cooperative or passive/parallel
MAR022 & MAR027 & MAR035	Dominant/passive	Cooperative or passive/parallel
MAR023 & MAR033	Cooperative or passive/parallel	Cooperative or passive/parallel
MAR024 & MAR028	Collaborative	Collaborative
MAR025 & MAR029	Collaborative	Collaborative
MAR030 & MAR031	Collaborative	Collaborative
MAR032 & MAR034	Cooperative or passive/parallel	Cooperative or passive/parallel

Note: Shading is used to indicate the changes in pair dynamics from Day 1 to Day 2.

Figure 15 Proportions of patterns of interaction in FFI+Collab on Day 1 and Day 2



As can be observed in the proportion of dyadic patterns (Figure 14 and Figure 15), the collaborative mode dominated in Collab and FFI+Collab on both DG days, although it did so to a greater extent in the former experimental condition. Moreover, the percentage of the collaborative pattern was identical on both days within each experimental condition (71% in Collab and 53% in FFI+Collab). Regarding Day 1, the second most frequent pattern in Collab was cooperative (or passive/parallel), representing 21% of the cases, while in FFI+Collab it was dominant/passive, accounting for 20% of the cases. Finally, in Collab, a single dyad was classified as dominant/passive, whereas in FFI+Collab, two dyads out of 15 were identified as cooperative, and another two as dominant/dominant. In what follows, examples are provided to illustrate these patterns.

(29) Collaborative pattern [ABE016 & ABE020 - Day 1]

*CHI B:	*aquí pon* 'Laura'… (put Laura here)
*CHI A:	*no*, the father, *los padres de Laura.* (Laura's parents)
*CHI B:	the Laura's father…
*CHI A:	*¿pongo esto?* (shall I put that?)
*CHI B:	no, Laura's father.
*CHI A:	[whispering] *eso*. (that's it)
*CHI B:	we prepare.
*CHI A:	no, their prepare.
*CHI B:	their prepare…
*CHI A:	sandwich.
*CHI B:	one sandwich with…
*CHI A:	*¿cómo se decía cacahuete?* (how did you peanut?)
*CHI B:	*eh… espera* (wait)*… eh…* [reads her notes] peanus&(peanuts) with…
*CHI A:	of…
*CHI B:	chocolate.
	[…]
*CHI B:	*¿qué hemos puesto?* (what did we put?) the Laura's father their prepare one sandwich with chocolate and peanuts.
*CHI A:	and one…
*CHI B:	and one apple.
*CHI A:	eh… Laura…
*CHI B:	'but', *pon* (put) 'but'.

In (29) we can see that CHI A and CHI B engage in a fluent dialogue where they both build up the ideas from the original text ("and one…", "and one apple"), and provide feedback to their peer's ("one sandwich with", "of") as well as their own suggestions ("no, Laura's father"). They consult each other ("*¿pongo esto?"* [shall I put this?], *"¿cómo se decía cacahuete?"* [how did you say peanut?]) and use the first person plural *("¿qué hemos puesto?"* [what did we put?]), indicating a feeling of shared ownership over the text.

(30) Cooperative (or passive/parallel) [ABE032 & ABE036 - Day 1]

*CHI B:	lunch time *o* (or) is the lunch time, *pon* (put) is the lunch time.
*CHI A:	*y* (and) in the afternoon is very hot.
*CHI B:	dad.
*CHI A:	dad?
*CHI B:	dad funny paces&(faces).
*CHI A:	*¿qué pone?* (what does it say?)
*CHI B:	dad funny [pronounced in Spanish] paces.
*CHI A:	eh… sister?
*CHI B:	*sí, pero pon primero* (yes but first put) in the afternoon (…) sister.
*CHI A:	sister?
*CHI B:	quickly to glasses, she is quicky to glasses.
*CHI A:	is wonderful day?
*CHI B:	[nods]

In (30), both children contribute to the writing ("in the afternoon is very hot", "dad funny paces"), but engage minimally with each other's suggestions, sometimes even just making use of gestures. Hence, the conversation takes the form of individual contributions rather than a true discussion or sharing of ideas on what they are writing. The only occasions when one of the learners asks something to his peer is not to question a particular linguistic choice, but rather to ask for a spelling clarification or content confirmation while writing (*"¿qué pone?"* [what does it say?], "eh… sister?"). The imperative mode predominates ("*pon primero* in the afternoon" [put first in the afternoon]).

(31) Dominant/dominant [MAR011 & MAR016 - Day 1]

*CHI B:	*¡que no empieza por* day! (I'm telling you it doesn't start with day!)
*CHI A:	*¡sí, porque* day *es hoy, hoy!* (yes, because day means today, today!)
*CHI B:	today!

*CHI A: *¿por qué?* (why?)
*CHI B: *que es así* (because it's like that).
*CHI A: *¡pues yo lo he escuchado!* (and I've heard it!)
*CHI B: day *es día* (…) *¿día Lola?* (day is day, day Lola?) [gesturing like it makes no sense].
[…]
*CHI B: priper&(prepare)!
*CHI A: *se pronuncia* prepar&(prepare) (it's pronounced prepar)
*CHI B: priper&(prepare)!
*CHI A: *¡que noooo! ¡que así no se dice!* (no way, it's not like that!)
*CHI B: *no, ya lo sé pero me refiero a la forma en escrito, nunca lo he visto así*.
(no, I know but I mean the written form, I've never seen it like that)
*CHI A: *¿entonces cómo se pone, lista?* (then how do you write it, smartass?)
*CHI B: *no sé* (I don't know).

In (31) children are constantly clashing over their suggestions (*"¿por qué?"* [why?], *"¡que es así!"* [it's like that!]). Their tone, very far from being constructive, even turns disrespectful sometimes (*"¿cómo se pone, lista?"* [how do you write it, smartass?]). In other words, these two learners seem to be keener to impose their own ideas rather than to convince their peer or find a compromise.

(32) Dominant/passive [ABE058 & ABE059 - Day 1]

*CHI A: [reading her notes] Laura goes to the, *o sea* (I mean), the father of Laura, eh… forgets prepare the sandwich of Laura, and sh… he gives money to buy a… no.
*CHI B: to go to the market.
*CHI A: to go to the market and a… a apple or something like that and Laura has, *bueno* (well), Laura buys a chocolate bar with peanuts and, *bueno* (well), she goes to the class and the teacher says she has eh… like a red…
*CHI B: *alergia* (allergy).
*CHI A: in the face they call to the… her father and her father tall&(talks) to the hospital.
*CHI B: hum.
*CHI A: and after go to class no she ate the chocolate bar is taste [pronounced in Spanish] amazing [pronounced correctly] amazing.
*CHI B: uhm?
*CHI A: the chocolate is taste amazing, *no sé, algo así decía, algo así,* the… *que estaba genial*. (I don't know, it said something like, something like, that it was great).
*CHI B: [whispering very low] xxx.

*CHI A:	*¿eh? bueno, pues pon* (huh? well, so then write) (…) go to the class to the class to classroom *o* (or)… (…) the teacher of Laura say she has a… the face with…
*CHI B:	with…
*CHI A:	red points? points?
*CHI B:	yes.

Finally, in (32), CHI A hogs the conversation, as can be clearly concluded from a simple look at the length of each learner's interventions. CHI A consults her peer only minimally ("red points? points?") and CHI B, who is in charge of writing, only participates in the dialogue with single words (*"alergia"* [allergy], "yes") and interjections ("hum"). The imperative mode appears sporadically ("*bueno, pues pon* go to the class" [well, so then write go to the class]).

With regards to the impact of TR on patterns of interaction, although, as mentioned earlier, the proportion of collaborative dyads in each experimental condition did not vary (and it continued to be the most frequent pattern), a closer look at the each of the pairs allows us to see some changes. In Collab, for instance, two dyads that were identified as collaborative on Day 1 decreased their mutuality and equality on Day 2. In the following example, an excerpt of the dialogue by ABE016 and ABE020 from Day 2 (who displayed a collaborative pattern on Day 1) is provided:

(33) Dominant/dominant [ABE016 & ABE020 - Day 2]

*CHI A:	in *no sé dónde xxx es muy caliente*. (in I don't know where it's very hot)
*CHI B:	*¿qué dices?* (what are you saying?)
*CHI A:	in afternoon he…
*CHI B:	*baina xxx el siguiente día*, in the afternoon is a hot? *¿es caliente?* (but the next day in the afternoon is a hot? is hot?)
*CHI A:	*pues sí, el siguiente es caliente.* (yes, the next is hot)
*CHI B:	*el siguiente día es muy caliente, pues no.* (the next day is very hot, I don't think so)
*CHI A:	*bueno, tú ponlo* (well, you write it down).
	[…]
*CHI B:	*pues ya está, a ver,* the grandfather… *no se dice* preparing, *eso lo tengo más claro que el agua,* the grandfather… (well, that's it, let's see, the grandfather… you don't say preparing, of that I'm totally sure, the grandfather…)
*CHI A:	prepare some ice cream.

*CHI B: ↑he prepare *pon* (put) he, *porfa* (please).
*CHI A: ¡que no! (no way!)
*CHI B: *bueno pues nada, no sé (…) bueno pues ya está, así.* (well, nothing then, I don't know, well that's it then)

In (33), these child learners cannot resolve their linguistic discussions in a constructive manner, and instead of contrasting or consulting their suggestions with their peer, as they had done in example (29), they try to impose their own choice. First, CHI A insists on the wrong expression "**is a hot*" even though her partner is not very convinced. In the second part, CHI B points at two different linguistic features: the verb form (prepare vs preparing) and the need for the subject pronoun 'he'. The former is not discussed with CHI A, who opts for the almost target-like form 'prepare'. Nonetheless, the second point of discussion generates another source of conflict, as CHI B ends up begging her peer to write 'he', but CHI A strongly refuses this choice and CHI B eventually gives up ("*bueno pues nada*" [well, nothing then]). The use of imperative verb forms also reinforces the notion of confrontation throughout their interaction (*"tú ponlo"* [you write it down], *"pon 'he', porfa"* [put 'he', please]).

However, two other dyads in Collab managed to increase their mutuality and equality and displayed a collaborative pattern on Day 2. That was the case of ABE058 & ABE059, as can be observed in the following example:

(34) Collaborative [ABE058 & ABE059 - Day 2]

*CHI A: *a ver* (let's see), the grandmother… *está jugando con esta* (is playing with this one) *y Tom está jugando con el tío* (and Tom is playing with his uncle), María is… *bueno* (well), the dad is laughing and, *bueno* (well), uncle and Tom there are playing football and is the mother and is taking photos and this is all.
*CHI B: *sí, pero yo creo que dice aquí coge la* portion *y se lo da a… a la tía porque está en la piscina.* (yes, but here I think it says that she takes a portion and that she gives it to her… aunt because she's at the swimming pool)
*CHI A: *umm vale* (alright), ok (…) *pues* (so) the Smiths.
: […]
*CHI B: *punto, ¿no?* (full stop, right?)
*CHI A: [taking the turn to write] *sí* (yes), ok eh… the grandmother…
*CHI B: is playing cards.
*CHI A: winning all the time.

*CHI B:	*no, '*old' *no* [spelling in Spanish] *a-l-l.*
*CHI A:	*¡ah, sí, claro!* (oh yes, sure)
	[…]
*CHI A:	[starts writing] afternoon or aftermoon?
*CHI B:	afternoon *ene* (en) [pronouncing in Spanish] noon.
*CHI A:	[exaggerating the double vowel] noon eh… and is preparing…
*CHI B:	preparing… some ice cream.
*CHI A:	[pronouncing in Spanish] preparing.
*CHI B:	*o* (or) ice cream.
*CHI A:	some ice cream [pronouncing in Spanish] some ice crem… *¡punto!* (full stop) eh… eh…

In (34), CHI B, who had remained silent in example (30) plays a more proactive role and speaks her mind throughout the dialogue, introducing nuances to CHI A's ideas (*"sí, pero yo cre*o…" [yes, but I think that]), suggesting the use of punctuation marks (*"punto, ¿no?"* [full stop, right?]), providing corrective feedback to her peer ("a-l-l") or to herself ("o ice cream" [or ice cream]). CHI A also consults CHI B and relies on her opinion to a larger extent (*"¡ah, sí, claro!"* [oh yes, sure], "afternoon *o* aftermoon?").

Regarding the impact of TR on FFI+Collab's patterns, all the dyads that were collaborative on the first day displayed the exact same pattern on Day 2. However, no other dyad with lower degrees of mutuality or equality managed to shift to a collaborative interaction on the second day. There were three dyads in which equality appeared to increase from Day 1 to Day 2, as they moved from a dominant/passive pattern to a more balanced cooperative interaction (yet with an equally low mutuality). Finally, MAR011 and MAR016, who had been identified as dominant/dominant in example (31), displayed a dominant/passive interaction on Day 2, hence showing a more unbalanced level of control over the task as compared to their first encounter with the collaborative DG task. The following excerpt illustrates their interaction on Day 2:

(35) Dominant/passive [MAR011 & MAR016 - Day 2]

*CHI A: *yo recuerdo que decía algo de la tarde y que como hacía tanto calor pues el* grandfather preparing the kitchen… *¿tú qué tienes?* [reads CHI B's notes] Tom is play pay football with uncle… Tom in afternoon is very hot… *ah pues perfecto, tú tienes esta palabra y yo tengo la palabra del* grandfather, *del abuelo* [CHI B does not react] *bueno, vamos a poner lo de Tom, ¿no?*
(I remember that it said something about the afternoon, and that, as it was very hot, the grandfather preparing the kitchen, what do you have? Tom is play football with uncle… Tom in afternoon is very hot… oh well perfect then, you have this word and I have the grandfather's word, the grandfather's, well let's write Tom's bit, right?)
[CHI B nods and CHI A starts writing]

*CHI A: Tom… *ah espera, que tú lo tienes aquí bien* (…) *luego dice algo de María y de su madre*… María is the mother… *tres puntos suspensivos, no sé, decía algo de su madre* (…) *pues si no lo sabemos pasamos,* in the afternoon very hot… *eso es muy poco, no vamos a poner por lo menos algo parecido:* María is the mother, *María está con su madre, ¿cómo se pone eso? (…) ay no, que María es con … ¡ya está!* Maria's mother.
(Tom… oh wait you have it correct here, then it says something about María and her mother… María is the mother… etcetera, I don't know, it said something about her mother, well I we don't know it we just move on, in the afternoon very)

*CHI B: *¿María es madre?* (María is mother?)

*CHI A: *no, ¿cómo se pone María está con su madre? ¿cómo se pone?*
(no, how do you say María is with her mother? how is it?)

*CHI B: María is…

*CHI A: the mother?

*CHI B: Maria is with her mother.

*CHI A: *vale, ahora voy a poner lo de*… [CHI B is writing more on her notes photocopy] *espera que voy a poner lo de*… in the… [CHI B gives her the photocopy] the afternoon… is… very… hot and… preparing… *¡y esto es lo único que nos sabemos!* RES [calls the researcher]!
(ok, now I'm going to put the part about… wait I'm going to put the… in the… the afternoon… is… very… hot and preparing… and this is all we know! RES! [calls the researcher])

In (35), we can clearly observe a dominant/passive pattern just from the turn distribution and length, which is dominated by CHI A. On this occasion, CHI B, instead of arguing with her partner, plays an almost silent role and lets CHI A make her own decisions. The only occasion on which CHI B reacts is when CHI A opts for "María's mother", which sounds wrong to her. After a short brainstorming ("María is…" "the mother?"), CHI B provides the target-like solution ("María is with her mother"). Although

CHI A does try to engage CHI B at several points in the dialogue, by directly calling her or using the first-person plural (*"¿tú qué tienes?"* [what do you have?], *"vamos a poner lo de Tom, ¿no?"* [let's write Tom's bit, right?] , *"¿cómo se pone?"* [how do you say?]), all these efforts are futile. After CHI A realizes it will only be her doing the task, she starts using the first person singular (*"ahora voy a poner"* [now I'm going to put], *"espera que voy a poner"* [wait, I'm going to put]), and only refers to themselves as a dyad when it becomes clear that the text is too short (*"¡y esto es todo lo que sabemos!"* [and this is all we know!]).

Turning now to RQ2b, we examined the texts produced in response to the DG task on Day 1 and Day 2 in the three experimental conditions (Comp, Collab, FFI+Collab). For the current analysis, the sample size in Comp is somewhat smaller (n = 26), as four participants did not complete the dictogloss task on Day 1 and other four were missing on Day 2. First, the results regarding clause type (Torras, 2005) are considered. The descriptive statistics are displayed in Table 33 (Day 1) and Table 34 (Day 2).

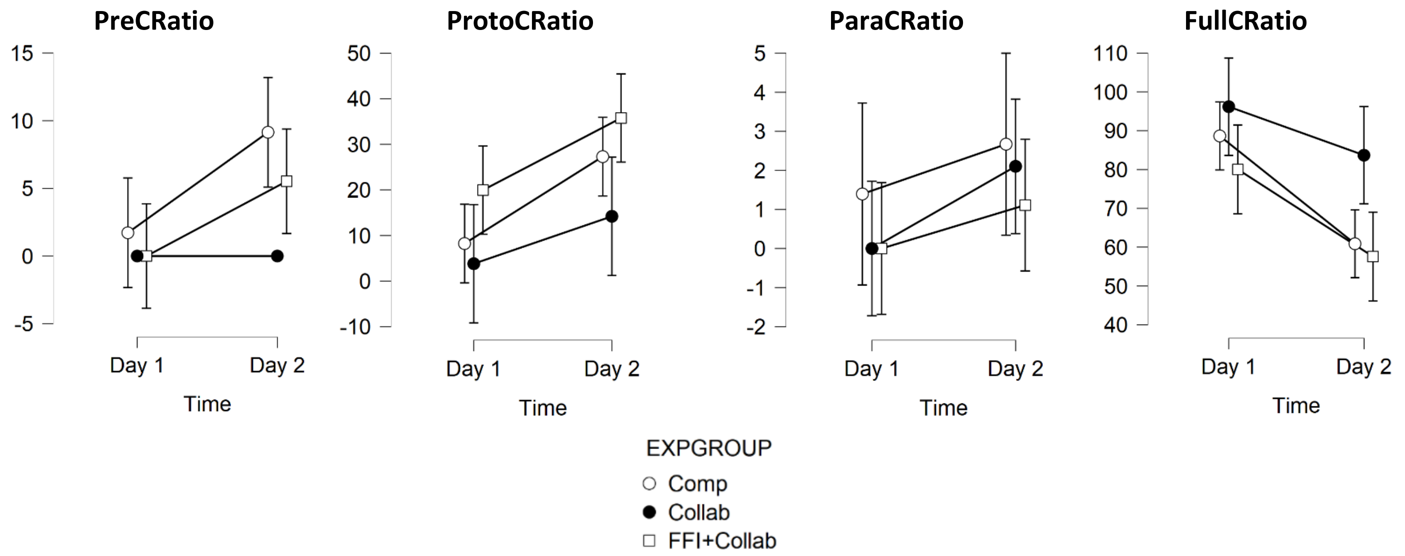
Table 33 Clause type ratio on Day 1 / experimental group

	Comp (n = 26)			Collab (n = 14)			FFI+Collab (n = 15)		
	M	**SD**	**95% CI**	**M**	**SD**	**95% CI**	**M**	**SD**	**95% CI**
PreCRatio	1.71	7.43	[-1.28 – 4.71]	0	0	-	0	0	-
ProtoCRatio	8.23	21.78	[-0.56 – 17.03]	3.81	10.03	[-1.98 – 9.61]	19.96	26.17	[5.47 – 34.45]
ParaCRatio	1.40	4.17	[-0.29 – 3.08]	0	0	-	0	0	-
FullCRatio	88.65	25.09	[78.52 – 98.79]	96.19	10.03	[90.40 – 101.98]	80.04	26.17	[65.55 – 94.53]

Table 34 Clause type ratio on Day 2 / experimental group

	Comp (n = 26)			Collab (n = 14)			FFI+Collab (n = 15)		
	M	**SD**	**95% CI**	**M**	**SD**	**95% CI**	**M**	**SD**	**95% CI**
PreCRatio	9.14	18.61	[1.82 – 16.47]	0	0	-	5.52	9.86	[0.61 – 10.99]
ProtoCRatio	27.30	30.35	[15.04 – 39.56]	14.21	30.19	[-3.22 - 31.64]	35.78	34.86	[16.48 – 55.09]
ParaCRatio	2.67	7.92	[-0.52 – 5.87]	2.10	4.21	[-0.33 - 4.54]	1.11	4.30	[-1.27 – 3.49]
FullCRatio	60.88	38.68	[45.26 – 76.51]	83.69	29.73	[66.52 - 100.85]	57.58	39.51	[35.70 – 79.46]

Figure 16 Interaction plots for type of clause. Error bars show 95% confidence intervals



As can be observed, on Day 1, learners from all three experimental conditions produced on average a larger proportion of full-fledged clauses than preclauses, protoclauses and paraclauses, implying that child learners on their first DG encounter were generally capable of producing clauses that were both grammatical and contextualized with the task. The second highest mean rate of clause type corresponded to protoclauses (i.e. contextualized but not grammatical clauses), although the average rate largely varied across groups, as it almost reached 20% in FFI+Collab (the highest mean rate) and 4% in Collab (the lowest mean rate). In contrast, preclauses (not contextualized and ungrammatical clauses) and paraclauses (uncontextualized but grammatical clauses) were only present in Comp, with very low mean rates (below 2).

On Day 2, however, there were some differences. Although, on average, the FullCRatio was again the highest of all four categories, there was a decrease as compared to Day 1 in the three experimental conditions, with the difference being sharper in Comp and FFI+Collab. Conversely, the ProtoCRatio increased in the three groups, but it was clearer in the case of FFI+Collab. Finally, the average of ParaCRatio continued to be very low, and that of PreCRatio increased in Comp and FFI+Collab, whereas Collab did not produce any instances of this type of clause. A series of repeated measures ANOVA were run in order to ascertain whether there were any significant differences (see APPENDIX C for the complete statistical output). The interaction plots can be found in Figure 16. For PreCRatio, ProtoCRatio and FullCRatio a significant main effect of time (with a small effect size) was found, respectively: $F = 7.74$, $p = .07$, $\eta^2 = 0.04$; $F = 13.59$, $p < .001$, $\eta^2 = 0.07$; $F = 24.33$, $p < .001$, $\eta^2 = 0.10$. Hence, learners' PreCRatio and ProtoCRatio was significantly higher on Day 2 than on Day 1 (t = -2.78, $p = .007$, CI [-1.20 – 1.55]; t = -3.69, $p < .001$, CI [-23.31 – -6.88]), whereas the average of FullCRatio was significantly lower (t = 4.93, $p < .001$, CI [12.40 – 29.41]).

With regards to text-based complexity and accuracy measures, Table 35 and Table 36 display the descriptive statistics for Day 1 and Day 2 in each of the experimental conditions.

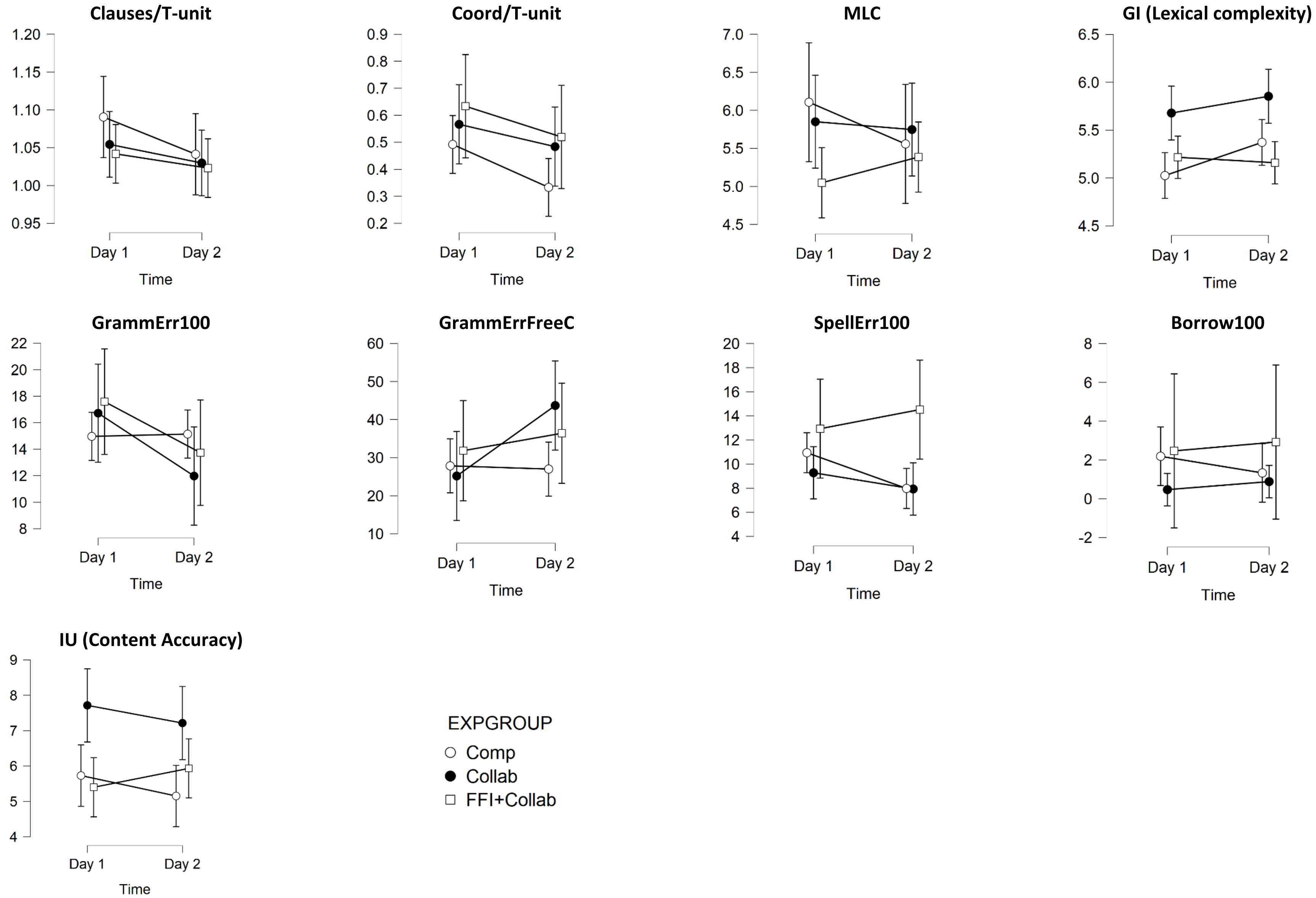
Table 35 Complexity (grammatical and lexical) and accuracy (general and content) on Day 1 / experimental group

	Comp (n = 26)			**Collab (n = 14)**			**FFI+Collab (n = 15)**		
	M	**SD**	**95% CI**	**M**	**SD**	**95% CI**	**M**	**SD**	**95% CI**
C/T-unit	1.09	0.16	[1.02 – 1.15]	1.05	0.08	[1 – 1.1]	1.04	0.09	[0.99 – 1.09]
CoordC/T-unit	0.49	0.33	[0.36 – 0.62]	0.57	0.31	[0.39 – 0.74]	0.63	0.29	[0.47 – 0.79]
MLC	6.11	2.74	[5 – 7.21]	5.85	1.07	[5.23 – 6.46]	5.05	0.76	[4.63 – 5.47]
GI	5.02	1.32	[4.49 – 5.56]	5.68	0.86	[5.18 – 6.17]	5.21	1.1	[4.61 – 5.82]
GramErr100	14.97	7.09	[12.10 – 17.83]	16.72	7	[12.67 – 20.76]	17.59	9.73	[12.20 – 22.98]
GramErrFreeC	27.85	25.72	[17.46 – 38.24]	25.19	20.11	[13.59 – 36.80]	31.83	28.51	[16.04 – 47.62]
SpellErr100	10.94	8.43	[7.53 – 4.34]	9.28	5.17	[6.29 – 12.26]	12.93	5.14	[10.09 – 15.78]
Borrow100	2.19	5.33	[0.03 – 4.34]	0.47	1.22	[-0.24 – 1.18]	2.47	7.01	[-1.41 – 6.35]
IU	5.73	3.58	[4.28 – 7.18]	7.71	2.20	[6.44 – 8.98]	5.40	3.09	[3.69 – 7.11]

Table 36 Complexity (grammatical and lexical) and accuracy (general and content) on Day 2 / experimental group

	Comp (n = 26)			**Collab (n = 14)**			**FFI+Collab (n = 15)**		
	M	**SD**	**95% CI**	**M**	**SD**	**95% CI**	**M**	**SD**	**95% CI**
C/T-unit	1.04	0.09	[1 – 1.08]	1.03	0.05	[1 - 1.06]	1.02	0.05	[0.99 – 1.05]
CoordC/T-unit	0.33	0.27	[0.22 – 0.44]	0.48	0.25	[0.33 - 0.63]	0.52	0.27	[0.37 – 0.67]
MLC	5.56	0.88	[5.20 – 5.91]	5.75	1.22	[5.04 - 6.45]	5.39	1.25	[4.69 – 6.08]
GI	5.37	1.09	[4.93 – 5.81]	5.85	0.75	[5.42 - 6.29]	5.16	1.12	[4.53 – 5.78]
GramErr100	15.14	6.80	[12.39 – 17.89]	11.98	6.15	[8.43 - 15.53]	13.74	8.84	[8.84 – 18.63]
GramErrFreeC	27	23.75	[17.41 – 36.60]	43.69	21.81	[31.09 - 56.28]	36.40	29.84	[19.88 – 52.93]
SpellErr100	7.98	8.15	[4.68 – 11.27]	7.93	7.65	[3.52-12.35]	14.52	11.22	[8.30 – 20.74]
Borrow100	1.33	2.60	[0.28 – 2.38]	0.88	1.61	[-0.05-1.81]	2.92	7.47	[-1.21 – 7.06]
IU	5.15	3.43	[3.77 – 6.54]	7.21	2.49	[5.78-8.65]	5.93	2.58	[4.51 – 7.36]

Figure 17 Interaction plots of the repeated-measures ANOVA for the grammatical and lexical complexity and general accuracy measures. Error bars show 95% confidence intervals



Starting with grammatical complexity, on Day 1, learners of all three experimental conditions did not use much subordination (C/T-unit), as there was almost a total correspondence between clauses and T-units (i.e. the values were close to 1). This is in line with the original dictogloss text subordination measures (DG1 = 1.05, DG2 = 1.12). In contrast, coordination (CoordC/T-unit) was much more present in their writings, as slightly more than half of the T-units fell under this category (the highest mean value corresponding to that of FFI+Collab). Moreover, the mean values in Collab and FFI+Collab were above the Coord/T-unit ratio found in the original dictogloss texts (DG1 = 0.55, DG2 = 0.44). Finally, the length-based grammatical complexity measure (MLC) shows that, on average, children used 5-6 words per clause on Day 1, somewhat below the original text clause length (DG1 = 6.42, DG2 = 7.05).

Looking at the same measurements on Day 2, we can observe that C/T-unit and MLC average was practically identical, whereas CoordC/T-unit showed a decrease. In fact, the repeated measures ANOVA indicated a significant main effect of time in CoordC/T-unit, with a small effect size ($F = 4.37$; $p = .042$, $\eta^2 = 0.04$). Moreover, the post-hoc test (with Bonferroni correction) showed that child learners, regardless of their experimental condition, used significantly fewer coordinated clauses per T-unit on Day 2 than on Day 1 (t = 2.09, p = .042, CI [0.01 – 0.23]). The same omnibus statistical test indicated a significant main effect of experimental condition, with a small effect size ($F = 3.25$; $p = .047$, $\eta^2 = 0.06$), but the post-hoc analysis failed to show any significant differences between the three conditions. Looking at the mean difference confidence interval, Comp and FFI+Collab were closest to a significant difference as it barely crossed zero: [-0.33 – 0.01]. Even in this case, the effect size was small: $d = -0.33$. In C/T-unit and MLC, the omnibus tests failed to show any significant main effects or interactions.

Turning to lexical complexity (GI), on Day 1 and Day 2, child learners obtained on average very similar measures, below 6. If compared to the original texts (DG 1 = 6.52, DG2 = 7.09), children's lexical richness was below those measures. The repeated measures ANOVA failed to show any significant main effects or interactions.

Regarding general accuracy measures, starting with grammatical accuracy, on Day 1, learners made on average 15-18 errors per 100 words (the highest error rate corresponding to FFI+Collab and the lowest to Comp). Conversely, looking at the rate of grammar error free clauses on the same day (GramErrFreeC), it can be observed that FFI+Collab had the highest accuracy rate, while Collab and Comp performed similarly. On Day 2, GrammErr100 decreased only in Collab and FFI+Collab. The statistical analysis revealed a significant effect of time , with a small effect size ($F = 5.92$; $p = .018$, $\eta^2 = 0.03$). In fact, learners on average made significantly fewer grammar errors on Day 2 than on Day 1 ($t = 2.43$, $p = .018$, CI [0.49 – 5.12]). The rate of GramErrFreeC increased from Day 1 to Day 2, but especially in the case of Collab, whose writing obtained on average almost 20 points more on the second DG encounter. However, the analysis failed to show any significant main effects or interactions.

With regards to mechanical accuracy (SpellErr100), child learners on Day 1 made on average of 9-13 errors per one-hundred words (the highest spelling error rate corresponding to FFI+Collab). On Day 2, Comp and Collab decreased their average number of spelling errors, while FFI+Collab increased it. Yet, the statistical analysis failed to show any significant main effects or interactions. Looking at lexical accuracy (Borrow100), on average, on Day 1, learners in Collab made very little use of borrowings (not even 1 in 100 words), whereas Comp and, especially, FFI+Collab did so to a greater extent. On Day 2, their average rates were very similar, with FFI+Collab and Collab showing a slight increase and Comp a slight decrease. In fact, the statistical analysis did not report any significant main effects or interactions.

Table 35 and Table 36 also display the results of Content Accuracy in terms of the average number of Idea Units (IUs) that child learners retrieved from the original DG texts. On Day 1, learners were able to recall in their writings an average of 5-8 IUs, with Collab showing the highest ratio (slightly above half of the original IU number) and FFI+Collab the lowest. On Day 2, the mean IU rate was very similar in all three conditions. The repeated-measures ANOVA failed to show any significant main effects or interactions.

Moving on the specific grammar accuracy measures, Table 37 and Table 38 display the descriptive statistics for 3S in DG1 and POSS and DG2, respectively. As the original tasks were presented in a counterbalanced order, the analysis of task repetition (i.e. time) was not considered relevant in this case. Instead, the analysis was based on the DG task target feature, regardless of the day on which learners carried it out. The *n* in each of the experimental conditions was obtained from those participants whose writing contained at least one 3S obligatory context in DG1 or one POSS obligatory context in DG2.

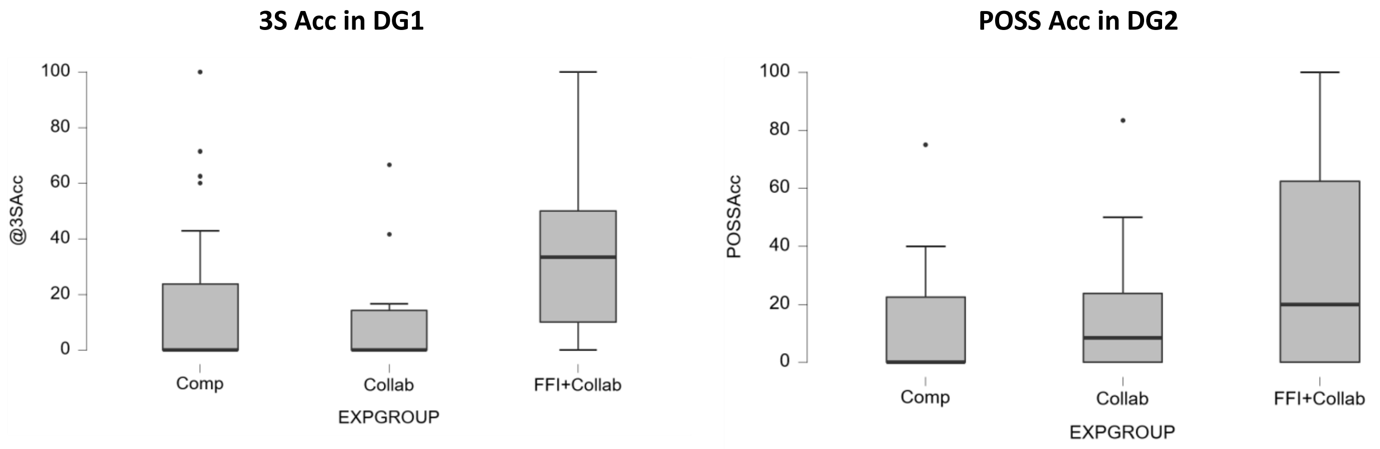
Table 37 Descriptive statistics for 3S in DG1

	Obligatory contexts			Correctly produced			Overproduction			3S Mean Accuracy Rate (3SAcc)		
	N	M	SD	N	M	SD	N	M	SD	M	SD	95% CI
Comp (n = 26)	145	5.58	2.77	36	1.38	2.62	2	0.08	.39	20.80	31.72	[7.99 – 33.61]
Collab (n = 13)	84	6.46	2.37	9	0.69	1.44	0	0	0	10.71	20.76	[-1.83 – 23.26]
FFI+Collab (n = 11)	49	4.45	1.92	23	2.09	1.87	6	0.55	1.04	35.78	30.69	[15.16 – 56.40]

Table 38 Descriptive statistics for POSS in DG2

	Obligatory contexts			Correctly produced			Overproduction			POSS Mean Accuracy Rate (POSSAcc)		
	N	M	SD	N	M	SD	N	M	SD	M	SD	95% CI
Comp (n = 27)	97	3.59	1.67	14	0.52	1.22	2	0.07	0.27	10.27	18.49	[2.93 – 17.56]
Collab (n = 14)	53	3.79	1.48	14	1	1.47	4	0.29	0.61	18.93	25.77	[4.05 – 33.81]
FFI+Collab (n = 12)	31	2.58	1.44	16	1.33	1.61	1	0.08	0.29	36.67	43.13	[9.26 – 64.07]

Figure 18 Box plots for 3S Acc in DG1 and POSS Acc in DG2 / experimental group



With regards to 3S, we can observe that, on average, learners in Collab produced the highest obligatory context rate, followed by Comp and FFI+Collab. In the case of correctly produced instances, FFI+Collab had the highest rate, followed by Collab and Comp. Overproduction of 3S was practically nonexistent in children's text. The 3S accuracy rate in obligatory contexts was then calculated on the basis of these three aforementioned values (Pica, 1983). The highest mean accuracy rate was that of FFI+Collab, followed by Comp and Collab. In any case, children's 3SAcc was far from 50%. A one-way ANOVA was conducted to determine if there were any significant differences between the three experimental conditions on 3SAcc, but it failed to show any significant effect.

As far as the descriptive statistics of POSS are concerned (Table 38), we can observe that, on average, learners in Collab generated again most obligatory contexts, followed by Comp and FFI+Collab. The highest correctly produced average corresponded to FFI+Collab, followed by Collab and Comp. The number of overproduction instances was very low across the three groups. In addition, the accuracy rate (POSSAcc) was again highest for FFI+Collab, although it was also the one that showed largest variation, as can be observed in the box plot in Figure 18. Moreover, in the same way as in 3SAcc, the mean accuracy rate did not reach 50% in any of the groups. The one-way ANOVA with Brown-Forsythe homogeneity correction failed to show any significant effect of experimental condition.

To conclude RQ2b, the results regarding the analytic rubric on Day 1 and Day 2 are summarized in Table 39 and Table 40, respectively. The interaction plots resulting from the repeated-measures analyses can be found in Figure 19:

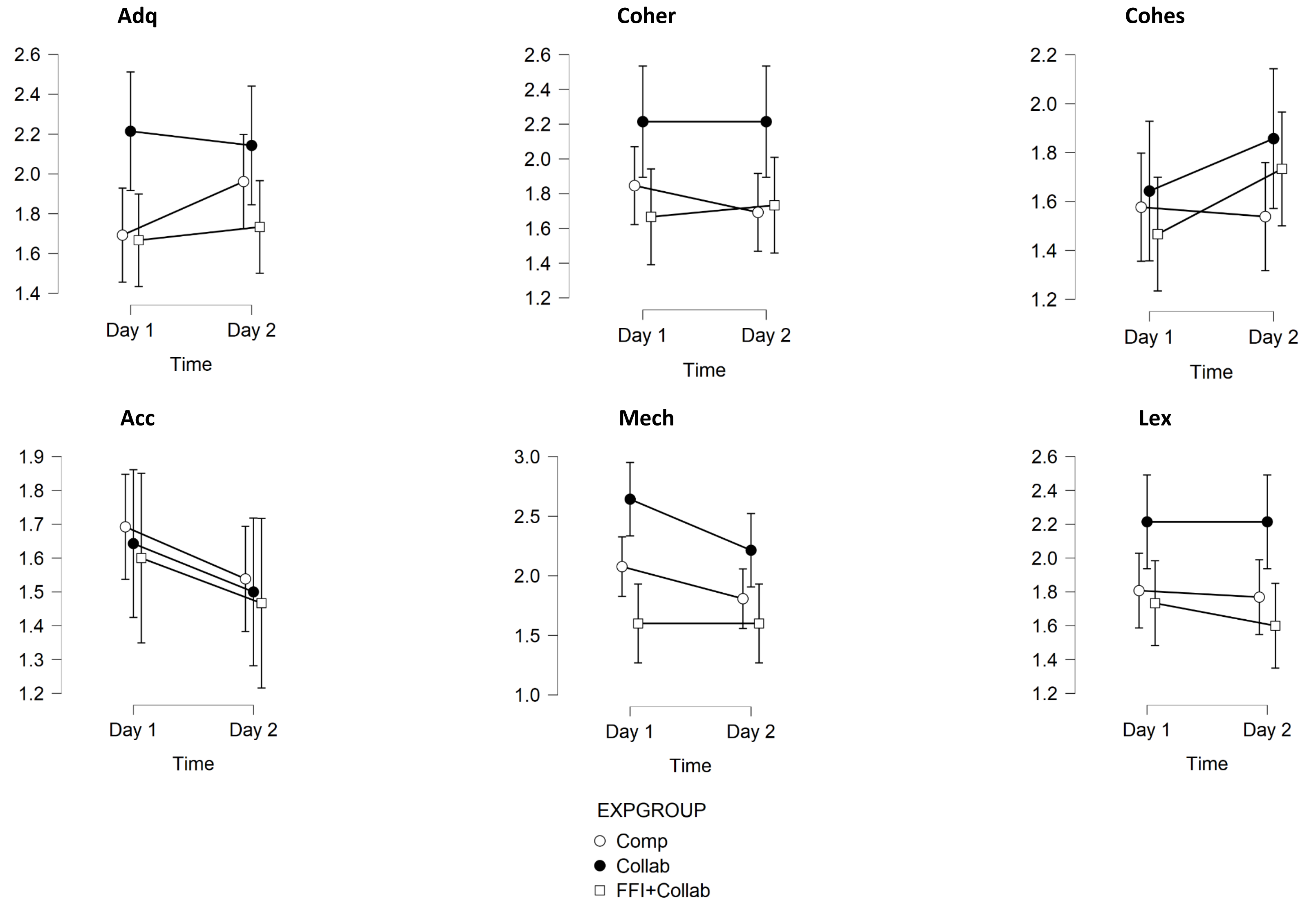
Table 39 Rubric measures on Day 1 / experimental group

	Comp (n = 26)			**Collab (n = 14)**			**FFI+Collab (n = 15)**		
	M	**SD**	**95% CI**	**M**	**SD**	**95% CI**	**M**	**SD**	**95% CI**
Adq	1.67	0.84	[1.35 – 1.98]	2.21	0.80	[1.75-2.68]	1.67	0.82	[1.21 – 2.12]
Coher	1.85	0.78	[1.53 – 2.16]	2.21	0.80	[1.75-2.68]	1.67	0.82	[1.21 – 2.12]
Cohes	1.58	0.76	[1.27 – 1.88]	1.64	0.74	[1.21-2.07]	1.47	0.74	[1.06 – 1.88]
Acc	1.69	0.62	[1.44 – 1.94]	1.64	0.50	[1.36-1.93]	1.60	0.83	[1.14 – 2.06]
Mech	2.08	0.80	[1.76 – 2.40]	2.64	0.74	[2.21-3.07]	1.60	0.83	[1.14 – 2.06]
Lex	1.81	0.69	[1.53 – 2.09]	2.21	0.70	[1.81-2.62]	1.73	0.80	[1.29 – 2.18]

Table 40 Rubric measures on Day 2 / experimental group

	Comp (n = 26)			**Collab (n = 14)**			**FFI+Collab (n = 15)**		
	M	**SD**	**95% CI**	**M**	**SD**	**95% CI**	**M**	**SD**	**95% CI**
Adq	1.97	0.85	[1.65 – 2.28]	2.14	0.77	[1.70-2.59]	1.73	0.80	[1.29 – 2.18]
Coher	1.69	0.74	[1.40 – 1.99]	2.21	0.58	[1.88-2.55]	1.73	0.70	[1.34 – 2.12]
Cohes	1.54	0.76	[1.23 – 1.85]	1.86	0.66	[1.47-2.24]	1.73	0.80	[1.29 – 2.18]
Acc	1.54	0.58	[1.30 – 1.77]	1.50	0.52	[1.20-1.80]	1.47	0.74	[1.06 – 1.88]
Mech	1.81	0.75	[1.51 – 2.11]	2.21	0.80	[1.75-2.68]	1.60	0.83	[1.14 – 2.06]
Lex	1.77	0.76	[1.46 – 2.08]	2.21	0.70	[1.81-2.62]	1.60	0.91	[1.10 – 2.10]

Figure 19 Interaction plots of the repeated-measures ANOVA for rubric measures. Error bars show 95% confidence intervals



Regarding the rubric task dimensions (Adq and Coher), on Day 1, we can observe that child learners in Collab scored the highest mean rates. Moreover, on a 1 (lowest) to 3 (highest) scale, Collab scored on average above 2 in Adq and Coher, whereas Comp and FFI+Collab scored mean scores below that benchmark. With regards to the language dimensions (Cohes, Acc, Mech, Lex) on Day 1, Collab obtained on average the highest mean rates (above 2.2) in all dimensions except for Acc, where Comp had a slightly higher score. Moreover, in general, Cohes and Acc were the dimensions with the lowest mean scores across the three experimental groups (ranging between 1.5 and 1.7).

On Day 2, Collab, with slight variations from Day 1, continued to have the highest mean scores of all three conditions in Adq, Coher, Cohes, Mech and Lex. Conversely, Comp again obtained the highest mean rate in Acc. Moreover, in general, Cohes and Acc were the two dimensions where the lowest mean rates were obtained across the three conditions (ranging from 1.5 to 1.9). In fact, Acc was the only dimension where the mean scores of all three groups decreased as compared to Day 1. A repeated-measures ANOVA was performed for each of the rubric dimensions, and a significant main effect of experimental condition was found in Mech, with a medium effect size (F = 5.62; p = .006, η^2 = .13). The post-hoc analysis (with Bonferroni adjustment) showed that, regardless of the DG day, on average, Collab scored significantly higher in Mech than FFI+Collab (t = 3.34, p =.005, CI [0.83 – 0.25]). In the rest of dimensions, the analysis failed to show any significant main effects or interactions.

In summary, the results for RQ2b do not allow us to observe an advantage of the collaborative mode over individual writing in the DG tasks, as there were no significant differences between Comp and the two collaborative conditions (Collab and FFI+Collab). On the contrary, learners in Collab seem to perform at a higher level in most writing dimensions, and this advantage seems to be greater when compared to FFI+Collab's performance (e.g. the former were significantly better in the rubric dimension of lexis than the latter). TR did not appear to influence learners' DG writing to a large extent, but there were a few exceptions. First, CoordC/T-unit showed a decrease in the amount of coordination used by child learners from Day 1 to Day 2. Secondly, the GramErr100 indicated that learners' writing (specifically, that of Collab and FFI+Collab) became more accurate on Day 2. Yet, the classification of the type of clauses, in spite of indicating a predominance of full-fledged clauses on both DG days, reflected a significant decrease of FullCRatio and a significant increase of ProtoCRatio and PreCRatio, irrespective of learners' task condition. Finally, in terms of the impact of FFI, FFI+Collab obtained the highest target form accuracy rate (both in 3S and POSS), but it failed to be statistically different from that of Comp and Collab.

RQ2c is concerned with the impact of collaborative work on subsequent individual writing. In order to investigate this question, the individual texts that children produced as a response to the story continuation task were analyzed. Child learners were tested before (T1) and after (T2) the DG intervention, as well as two weeks after (T3). The sample corresponds to those participants who carried out the DG task on Day 1 and Day 2 and completed the writing tests at T1, T2 and T3.

Regarding the development of clause type, Table 41, Table 42 and Table 43 summarize the results from the pretest (T1), posttest (T2) and delayed posttest (T3), respectively, and Figure 20 displays the interaction plots.

Figure 20 Interaction plots of the repeated-measures ANOVA for clause type. Error bars show 95% confidence intervals

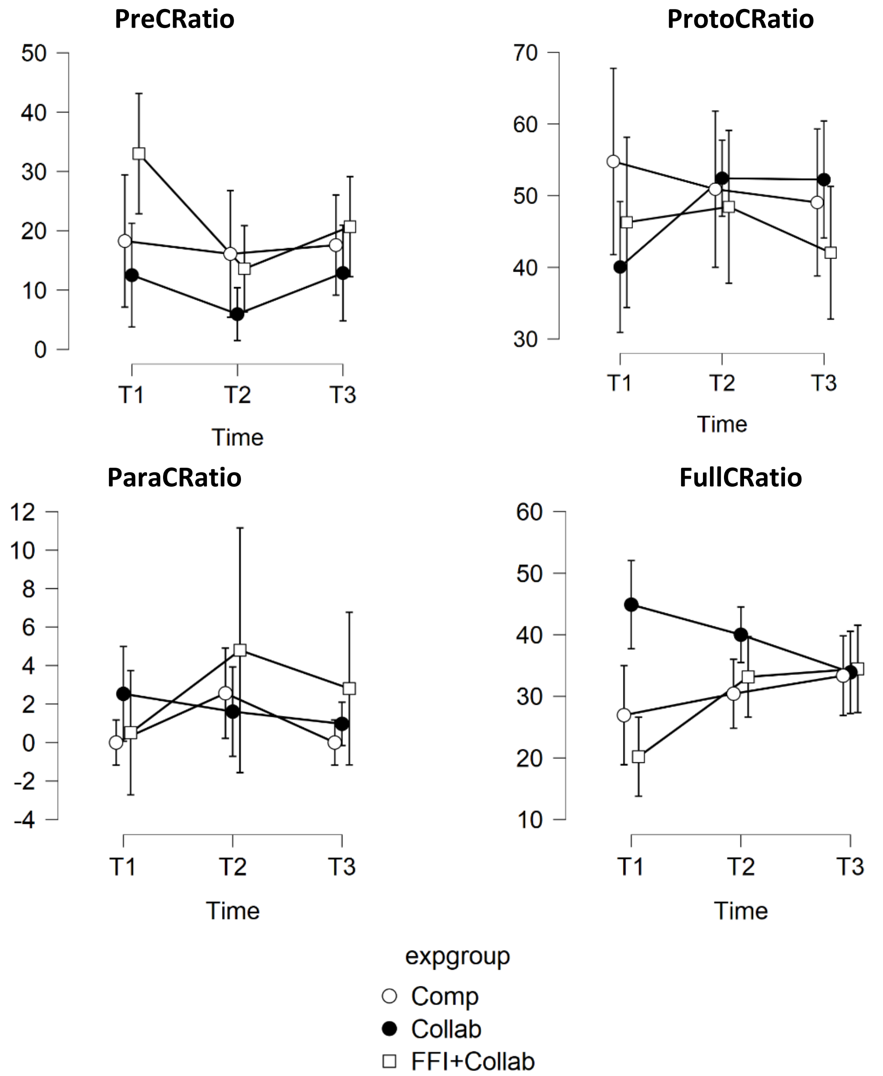


Table 41 Clause type ratio at T1 / experimental group

	Comp (n = 25)			Collab (n = 25)			FFI+Collab (n = 28)		
	M	**SD**	**95% CI**	**M**	**SD**	**95% CI**	**M**	**SD**	**95% CI**
PreCRatio	18.28	32.29	[4.95 – 31.60]	12.52	28.48	[0.76 – 24.28]	33.01	38.13	[18.22 – 47.80]
ProtoCRatio	54.78	33.74	[40.85 – 68.70]	40.05	32.61	[26.59 – 53.51]	46.28	35.19	[32.64 – 59.93]
ParaCRatio	0	0	-	2.53	7.48	[0.55 – 5.62]	0.51	2.70	[-0.54 – 1.56]
FullCRatio	26.95	32.35	[13.59 – 40.30]	44.89	35.54	[30.22 – 59.57]	20.19	34.48	[6.82 – 33.56]

Table 42 Clause type ratio at T2 / experimental group

	Comp (n = 25)			Collab (n = 25)			FFI+Collab (n = 28)		
	M	**SD**	**95% CI**	**M**	**SD**	**95% CI**	**M**	**SD**	**95% CI**
PreCRatio	16.10	25.51	[5.57 – 26.63]	5.95	12.54	[0.78 – 11.13]	13.59	25.11	[3.86 – 23.33]
ProtoCRatio	50.91	27.65	[39.49 – 62.32]	52.43	25.23	[42.02 – 62.84]	48.45	35.47	[34.69 – 62.20]
ParaCRatio	2.56	6.96	[-0.31 – 5.43]	1.61	4.49	[-0.25 – 3.46]	4.79	19.21	[-2.65 – 12.24]
FullCRatio	30.43	28.87	[18.51 – 42.34]	40.01	29.11	[27.99 – 52.02]	33.16	37.94	[18.45 – 47.88]

Table 43 Clause type ratio at T3 / experimental group

	Comp (n = 25)			Collab (n = 25)			FFI+Collab (n = 28)		
	M	**SD**	**95% CI**	**M**	**SD**	**95% CI**	**M**	**SD**	**95% CI**
PreCRatio	17.59	22.43	[8.33 – 26.84]	12.88	27.46	[1.54 – 24.21]	20.69	29.98	[9.06 – 32.32]
ProtoCRatio	49.05	28.32	[37.37 – 60.74]	52.25	27.92	[40.72 – 63.77]	42.03	27.85	[31.23 – 52.83]
ParaCRatio	0	0	-	0.97	3.42	[-0.44 – 2.38]	2.81	9.22	[-0.77 – 6.38]
FullCRatio	33.36	35	[18.91 – 47.81]	33.90	27.04	[22.74 – 45.06]	34.47	31.85	[22.12 – 46.82]

At T1, learners in Collab produced the most appropriate writing, as they had the highest mean rate in FullCRatio and the lowest PreCRatio. Conversely, FFI+Collab produced on average more than twice as many preclauses than Collab and less than half full-fledged clauses. The average ProtoCRatio remained very similar across the three conditions, being Comp the group that yielded the highest mean rate, and ParaCRatio was very low (on average, less than 3% of all clauses).

Looking at T2, Collab was still the group whose writing was most appropriate, as it reflected again the highest FullCRatio and the lowest PreCRatio. Yet, the differences between Collab and FFI+Collab diminished with regards to FullCRatio, as the average difference between the two groups was less than 10 points. The average PreCRatio decreased in all three conditions as compared to T1, while ProtoCRatio slightly increased in Collab and Comp. The average ParaCRatio increased in the three groups, but it was still very low, representing on average less than 5% of all clauses.

Finally, at T3, the average FullCRatio, indicative of grammatical and functional appropriateness, was most similar across the three conditions than at T1 and T2. In fact, Comp and FFI+Collab continued to increase their average full-fledged clause ratio with respect to T1 and T2, whereas Collab had a downward trend from T1 to T3, as can be seen in the interaction plot. Moreover, at T3, it was FFI+Collab the group that obtained the highest FullCRatio average. The three conditions also became more similar in their PreCRatio average, where FFI+Collab had the highest mean production of this type of ungrammatical and functionally inappropriate clauses. In the same way as at T1, the average ProtoCRatio and ParaCRatio was very similar across the three groups. In sum, regardless of the testing time, the protoclause mean ratio was the highest of all four clause categories, accounting on average for almost 40%-55% of the clauses produced by child learners in each of the groups.

A series of repeated-measures ANOVA were run in order to check for the influence of testing time and experimental condition in each of the clause types (the complete statistical output can be found in C2, in APPENDIX C). In the case of PreCRatio, a significant main effect of Time was found, with a small effect size: $F = 3.61$, $p = .029$, $\eta^2 = 0.02$. The post-hoc analysis with Bonferroni correction showed that there was a significant difference between the average of preclauses produced at T1 and at T2: t =

2.68, p = .024, CI [0.92 – 17.85]. That is, on average, regardless of the experimental condition, PreCRatio was significantly lower at T2 than at T1. Next, in the case of FullCRatio a significant interaction was found between testing time and experimental condition, with a small effect size: F = 4.42, p = .002, η^2 = 0.02. The post-hoc analysis revealed that FFI+Collab had a significantly higher FullCRatio at T3 than at T1: t = 3.27, p = .047, CI [0.07 - 28.47]. We failed to find other significant main effects or interactions in the rest of the analyses.

With regards to the grammar and lexical complexity and accuracy measures, Table 44, Table 45 and Table 46 summarize the results for each testing time, and Figure 21 displays the interaction plots. Looking at the first grammar interdependence measure (C/T-unit), we can observe that on average child learners did not make a great use of subordination in their writing (i.e. T-units approximately equaled the number of clauses), regardless of the testing time or experimental grouping condition. In contrast, Coord/T-unit average values indicate that these learners often resorted to coordination, representing at least half of the T-units in their writing. While Comp and FFI+Collab yielded a very similar Coord/T-unit at three testing times, Collab lowered their ratio from T1 to T2 (almost 10 tenths), but they raised it again at T3 to a similar level of that at T1. In other words, Collab produced on average the highest Coord/T-unit of all three conditions at T1, the lowest at T2, and again the highest at T3. The repeated-measures ANOVA failed to show any significant main effects or interactions in either C/T-unit or Coord/T-unit.

Regarding the length-based grammar complexity measure (MLC), once again the values in general remained very similar across the three conditions, although there were some differences in the trends. Although the average number of words per clauses decreased from T1 to T2, it was more evident in Comp, as they did so by slightly more than a word. MLC did not vary in Collab across the three testing times (using approximately 6 words per clause on every occasion). In FFI+Collab MLC displayed a downward trend, and these learners produced on average a word less from T1 to T3. However, no significant main effects or interactions were found.

Turning to lexical complexity (GI), in contrast to what has been described about grammatical complexity, this lexical diversity index indicated wide differences between the three experimental conditions across the three testing times. Thus, Collab had the highest average values at T1, T2 and T3, followed by Comp and FFI+Collab. Although both Comp and FFI+Collab increased slightly their lexical diversity from T1 to T2, only Comp could maintain it at T3, as FFI+Collab showed their lowest GI mean value at T3. The repeated-measures ANOVA revealed that, in fact, experimental condition played a significant role in GI, with a medium effect size: $F = 9.83$, $p < .001$, $\eta^2 = 0.17$. The post-hoc test indicated that, regardless of the testing time, Collab was significantly different from FFI+Collab: $t = 4.41$, $p < .001$, CI [0.42 – 1.42].

The results of general accuracy measures are now described. First, GramErr100 shows that, on average, Collab produced fewer errors than Comp and FFI+Collab at every testing time, and in fact, Collab displayed very little variation across time (ranging from 15-16 grammar errors per 100 words). Conversely, Comp progressively produced more errors in their writing from T1 to T3, whereas FFI+Collab had a clear decrease from T1 to T2, but errors increased again from T2 to T3. Regarding the grammar error free clause ratio (GramErrFreeC), a similar pattern can be perceived. Collab produced the highest mean percentage of accurate clauses at T1, T2 and T3 (slightly higher than 30%), followed by Comp, both groups not showing much variation across time. Yet, FFI+Collab increased this rate from T1 to T2 by 10 points, and they were able to maintain it at T3, hence situating their written production closer to that of Comp and Collab. However, we failed to find any significant mains effects or interactions in neither of the general grammatical accuracy measures.

The rate of spelling errors per 100 words (SpellErr100) shows slight differences between the experimental conditions and testing times. In general, this ratio ranged between 6-11 errors. However, two different trends can be observed: while Comp and Collab decreased their average SpellErr100 from T1 to T2, the opposite was true for FFI+Collab. Finally, with regards to the lexical borrowing ratio (Borrow100), FFI+Collab produced on average the highest rate at the three testing times. In fact, although they did lower the number of borrowings from T1 to T2, they increased it again at T3. In the case of Comp and Collab, the number of borrowings went from being very low at T1 to becoming increasingly larger at T2 and T3, getting closer to the extensive use made by their counterparts in FFI+Collab. Nevertheless, the statistical analyses failed to show any significant main effects or interactions in either accuracy dimension.

Table 44 Complexity (grammatical and lexical) and general accuracy at T1 / experimental group

	Comp (n = 25)			Collab (n = 25)			FFI+Collab (n = 28)		
	M	**SD**	**95% CI**	**M**	**SD**	**95% CI**	**M**	**SD**	**95% CI**
C/T-unit	1.11	0.16	[1.04 - 1.17]	1.05	0.08	[1.02 - 1.09]	1.08	0.15	[1.03 – 1.15]
CoordC/T-unit	0.48	0.30	[0.34 – 0.61]	0.60	0.26	[0.48 – 0.70]	0.54	0.38	[0.38 – 0.69]
MLC	7.01	2.75	[5.81 – 8.20]	6.21	1.41	[5.63 – 6.79]	6.42	4.26	[4.70 – 8.14]
GI	3.99	1.09	[3.54 – 4.44]	4.48	0.66	[4.20 – 4.75]	3.58	0.94	[3.21 – 3.95]
GramErr100	15.61	7.55	[12.49 – 18.73]	15.11	7.21	[12.14 – 18.09]	19.90	11.98	[15.51 – 24.79]
GramErrFreeC	27.22	23.83	[16.91 – 37.52]	32.02	21.82	[23.01 – 41.03]	18.70	22.33	[9.67 – 27.72]
SpellErr100	10.69	19.76	[2.53 – 18.65]	8.45	7.24	[5.46 – 11.43]	8.41	10.79	[4.23 – 12.60]
Borrow100	2.10	4.26	[0.34 – 3.86]	1.59	3.02	[0.35 – 2.84]	7	9.41	[3.36 – 10.65]

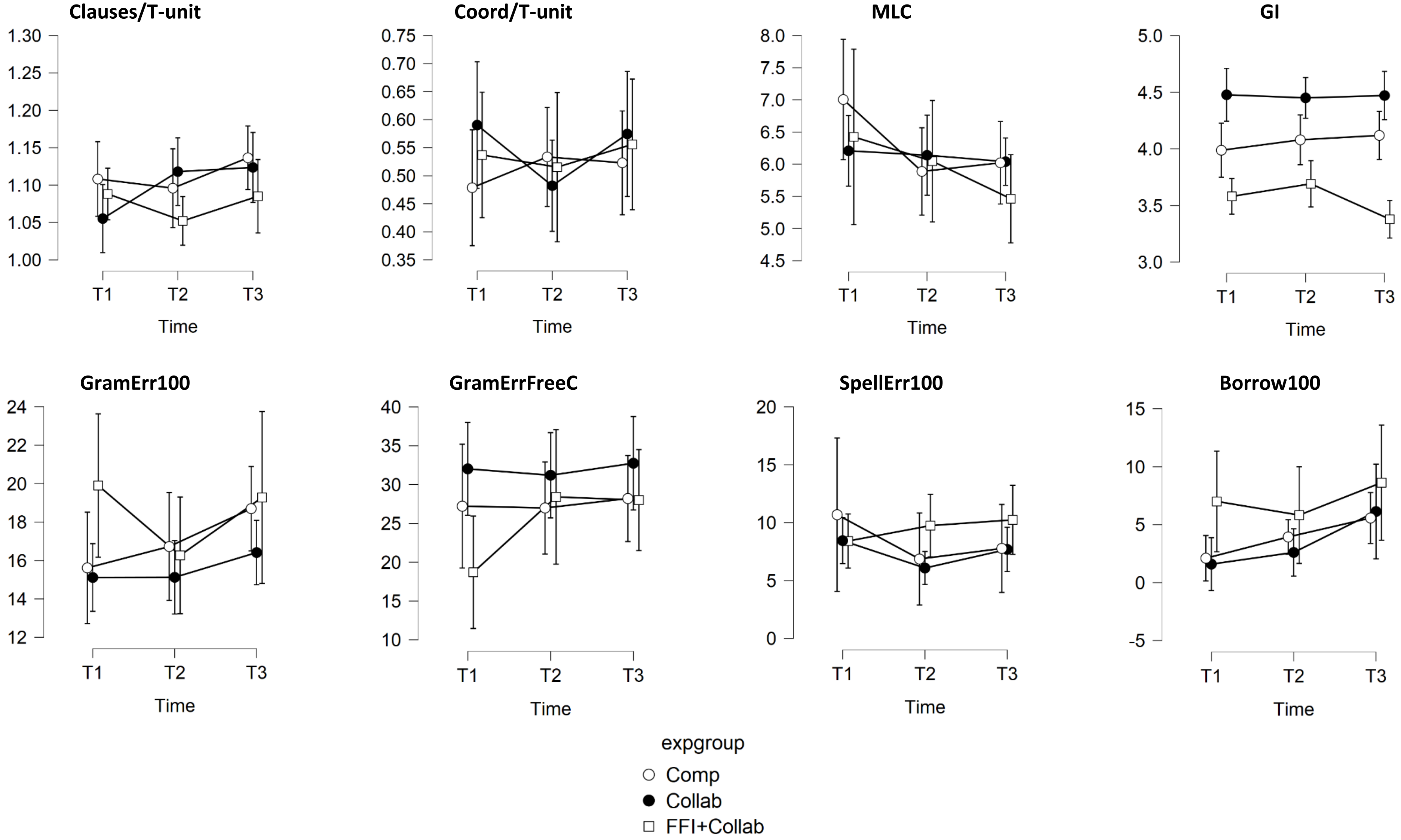
Table 45 Complexity (grammatical and lexical) and general accuracy at T2 / experimental group

	Comp (n = 25)			Collab (n = 25)			FFI+Collab (n = 28)		
	M	**SD**	**95% CI**	**M**	**SD**	**95% CI**	**M**	**SD**	**95% CI**
C/T-unit	1.09	0.18	[1.02 – 1.17]	1.12	0.13	[1.06 – 1.17]	1.05	0.09	[1.01 – 1.09]
CoordC/T-unit	0.53	0.29	[0.41 – 0.66]	0.48	0.27	[0.37 – 0.59]	0.51	0.37	[0.37 – 0.66]
MLC	5.89	1.37	[5.30 – 6.48]	6.14	1.74	[5.42 – 6.86]	6.05	1.55	[5.42 – 6.67]
GI	4.08	0.93	[3.69 – 4.47]	4.45	0.77	[4.13 – 4.77]	3.69	0.89	[3.35 – 4.04]
GramErr100	16.73	7.95	[13.44 – 20.01]	15.12	5.59	[12.81 – 17.43]	16.26	9.13	[12.72 – 19.81]
GramErrFreeC	26.98	21.55	[17.67 – 36.30]	31.20	17.92	[23.80 – 28.60]	28.40	26.96	[17.51 – 39.29]
SpellErr100	6.86	7.59	[3.73 – 9.99]	6.08	4.28	[4.31 – 4.84]	9.74	8.40	[6.49 – 13.01]
Borrow100	3.94	4.60	[2.04 – 5.84]	2.60	3.38	[1.21 – 4]	5.82	13.83	[0.46 – 11.18]

Table 46 Complexity (grammatical and lexical) and general accuracy at T3 / experimental group

	Comp (n = 25)			Collab (n = 25)			FFI+Collab (n = 28)		
	M	**SD**	**95% CI**	**M**	**SD**	**95% CI**	**M**	**SD**	**95% CI**
C/T-unit	1.14	0.19	[1.05 – 1.22]	1.12	0.13	[1.07 – 1.18]	1.08	0.14	[1.03 – 1.14]
CoordC/T-unit	0.52	0.27	[0.40 – 0.64]	0.57	0.25	[0.47 – 0.68]	0.55	0.31	[0.43 – 0.68]
MLC	6.02	0.94	[5.62 – 6.43]	6.04	1.58	[5.39 – 6.69]	5.46	1.24	[4.96 – 5.96]
GI	4.12	0.75	[3.81 – 4.43]	4.47	0.77	[4.15 – 4.79]	3.38	0.83	[3.05 – 3.70]
GramErr100	18.70	9.38	[14.82 – 22.57]	16.41	6.77	[13.62 – 19.21]	19.28	11.89	[14.66 – 23.98]
GramErrFreeC	28.20	21.96	[18.70 – 37.70]	32.75	21.73	[23.78 – 41.72]	28	25.40	[17.73 – 38.27]
SpellErr100	7.77	6.10	[5.25 – 10.29]	7.69	6.06	[5.19 – 10.19]	10.24	9.83	[6.43 – 14.06]
Borrow100	5.58	7.07	[2.66 – 8.49]	6.13	14.29	[0.24 – 12.03]	8.63	18.40	[1.49 – 15.76]

Figure 21 Interaction plots of the repeated-measures ANOVA for grammar and lexical complexity and general accuracy. Error bars show 95% confidence intervals



The analysis of the target form accuracy is now considered. For this purpose, the intergroup comparison at each time point will be initially considered (see the distribution of the results for 3S and POSS in Figure 22 and Figure 23, respectively). A series of one-way ANOVAs were conducted for each time to check for a main effect of experimental condition. The sample of participants at each testing time varies depending on the number of participants who produced obligatory contexts in their writing.

First, the 3S results corresponding to T1, T2 and T3 are reported (displayed in Table 47, Table 48 and Table 49, respectively). On the first testing day (T1), Collab produced on average the highest amount of obligatory 3S contexts, followed by Comp, and far behind, FFI+Collab. However, the highest mean accuracy rate corresponded to Comp, which was twice as high as that of their counterparts. At T2, it was Collab again that produced the highest average of obligatory 3S contexts (followed closely by the other two experimental conditions), but this time FFI+Collab displayed the highest accuracy rate. Finally, at T3, the pattern was similar to T2: Collab yielded on average most obligatory contexts, but FFI+Collab had the highest mean accuracy (with a large within group variation, as shown by SD values and the large box). What is evident from the accuracy scores distribution (Figure 22) is that the median score was very low at each testing time in all three conditions (below 20%), and that there were several outliers.

Regarding POSS, the results from T1, T2 and T3 are presented in Table 50, Table 51 and Table 52, respectively. At T1, among the learners who produced obligatory contexts in their writing, those in Comp had the highest mean, followed very closely by FFI+Collab and Collab. Moreover, Comp also had the highest accuracy mean rate, more than 10 points higher than the average of Collab and FFI+Collab, whose mean accuracy was very similar (around 30%). At T2, learners in Collab produced on average more obligatory contexts than Comp and FFI+Collab, and they also had the highest mean accuracy rate, reaching almost 50%. Finally, at T3, Collab generated on average more obligatory contexts than Comp and FFI+Collab, but this time it was the latter that obtained the highest accuracy rate, reaching almost 80%. Taken together, there was also a great variability within each of the groups at each testing time, as indicated by the large box lengths of the box plots (Figure 23).

Table 47 3S production results at T1

	Obligatory contexts			Correctly produced			Overproduction			3S Mean Accuracy Rate (3SAcc)		
	N	M	SD	N	M	SD	N	M	SD	M	SD	95% CI
Comp (n = 18)	72	4	2.11	21	1.17	1.46	8	0.44	0.92	31.22	38.38	[12.13 – 50.30]
Collab (n = 25)	146	5.84	3.89	18	0.72	1.17	2	0.08	0.28	12.84	24.32	[2.80 – 22.88]
FFI+Collab (n = 22)	61	2.77	2.54	15	0.68	1.67	1	0.04	0.21	11.59	24.56	[0.71 – 22.48]

Table 48 3S production results at T2

	Obligatory contexts			Correctly produced			Overproduction			3S Mean Accuracy Rate (3SAcc)		
	N	M	SD	N	M	SD	N	M	SD	M	SD	95% CI
Comp (n = 21)	101	4.81	3.53	22	1.05	2.42	2	0.10	0.30	15.70	26.26	[3.74 – 27.65]
Collab (n = 25)	136	5.44	4.23	13	0.52	1.16	2	0.08	0.28	6.04	12.17	[1.02 – 11.07]
FFI+Collab (n = 19)	78	4.11	3.23	30	1.58	3.34	1	0.05	0.23	19	16.66	[4.96 – 35.04]

Table 49 3S production results at T3

	Obligatory contexts			Correctly produced			Overproduction			3S Mean Accuracy Rate (3SAcc)		
	N	M	SD	N	M	SD	N	M	SD	M	SD	95% CI
Comp (n = 21)	95	4.52	2.98	14	0.67	1.52	4	0.19	0.40	11.76	25.01	[0.37 – 23.14]
Collab (n = 24)	131	5.46	2.84	16	0.67	1.52	1	0.04	0.20	9.56	19.79	[1.20 – 17.91]
FFI+Collab (n = 22)	79	3.59	2.84	22	1	2.29	0	0	0	23.11	38.31	[6.11 – 40.09]

Figure 22 Box plots for 3S Acc at T1, T2 and T3

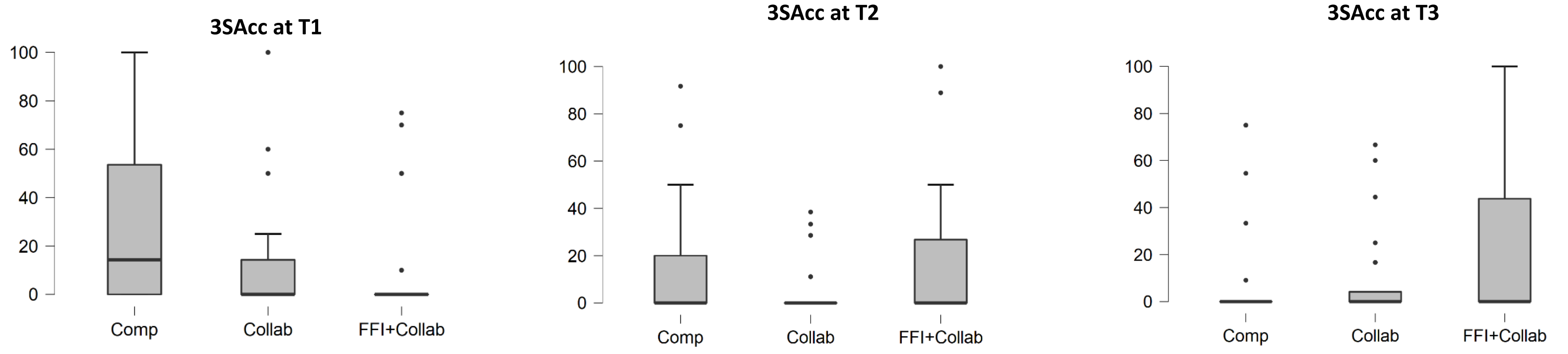


Table 50 POSS production results at T1

	Obligatory contexts			Correctly produced			Overproduction			POSS Mean Accuracy Rate (POSSAcc)		
	N	M	SD	N	M	SD	N	M	SD	M	SD	95% CI
Comp (n = 9)	15	1.67	0.71	7	0.78	0.67	1	0.11	0.33	41.67	40.40	[10.61 – 72.72]
Collab (n = 15)	24	1.60	1.35	7	0.47	0.52	2	0.13	0.52	28.20	42.70	[2.40 – 54.01]
FFI+Collab (n = 14)	23	1.64	1.08	9	0.64	0.84	1	0.07	0.27	29.17	40.28	[3.57 – 54.76]

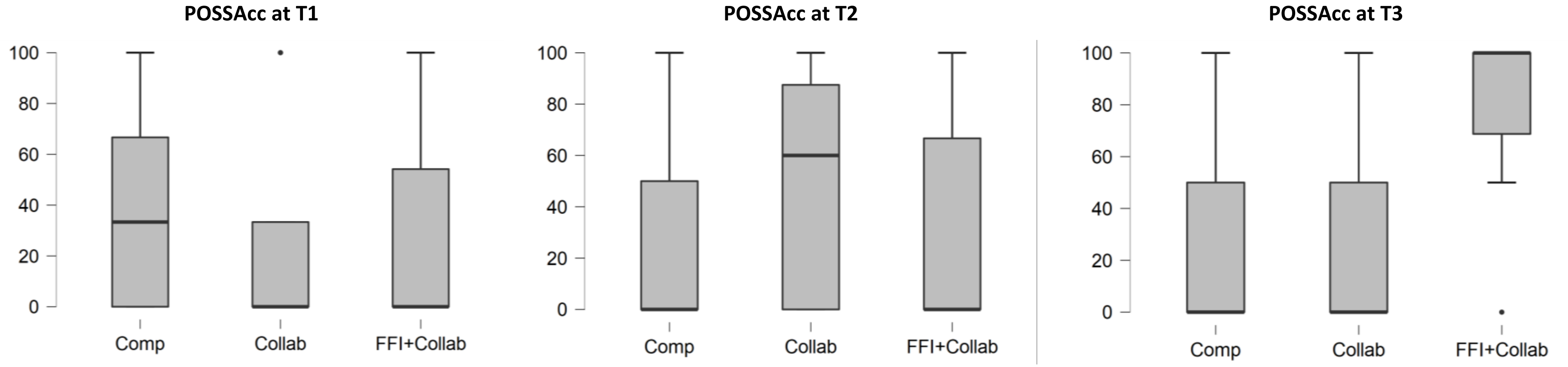
Table 51 POSS production results at T2

	Obligatory contexts			Correctly produced			Overproduction			POSS Mean Accuracy Rate (POSSAcc)		
	N	M	SD	N	M	SD	N	M	SD	M	SD	95% CI
Comp (n = 11)	20	1.82	1.40	6	0.55	0.93	0	0	0	27.27	41.01	[-0.28 – 54.82]
Collab (n = 15)	40	2.67	1.40	23	1.53	1.41	2	0.13	0.52	49.11	41.59	[26.08 – 72.14]
FFI+Collab (n = 14)	30	2.14	1.41	11	0.79	1.12	0	0	0	33.59	43.47	[7.32 – 59.86]

Table 52 POSS production results at T3

	Obligatory contexts			Correctly produced			Overproduction			POSS Mean Accuracy Rate (POSSAcc)		
	N	M	SD	N	M	SD	N	M	SD	M	SD	95% CI
Comp (n = 9)	15	1.67	1.32	5	0.56	1.13	3	0.33	0.71	28.57	48.79	[-16.56 – 73.70]
Collab (n = 13)	28	2.15	1.46	5	0.38	0.65	2	0.15	0.55	25	39.89	[-0.34 – 50.34]
FFI+Collab (n = 9)	15	1.67	1.12	12	1.33	1	1	0.11	0.33	78.12	36.44	[47.66 – 108.59]

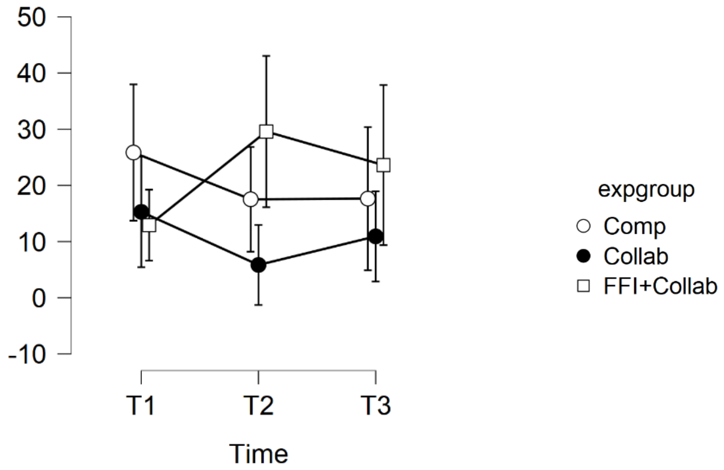
Figure 23 Box plots for POSS Acc at T1, T2 and T3



If the two target forms are compared, we can observe that, in general, the writing tests triggered more 3S than POSS production. A repeated-measures ANOVA was conducted at each testing time taking the two target features as the within-subjects independent variable, the experimental condition as the between-subjects factor and the obligatory contexts as the dependent variable. At T1, there was a significant interaction between target feature and experimental condition, with small effect size: $F = 3.81$, $p = .033$, $\eta^2 = 0.06$. The post-hoc analysis showed that the difference between the mean obligatory 3S and POSS contexts was significant for Collab: t = 4.60, $p < .001$, CI [1.74 – 8.71]. At T2, a significant main effect of target feature was found, with a medium effect size: $F = 25.72$, $p < .001$, $\eta^2 = 0.24$. In fact, child learners, regardless of their experimental condition, produced significantly more 3S than POSS obligatory contexts: t = 5.07, $p < .001$, CI [2.09 – 4.89]. Finally, at T3, another main effect of target feature was found, with a large effect size: $F = 50.19$, $p < .001$, $\eta^2 = 0.39$. Again, learners produced significantly more 3S than POSS obligatory contexts: t = 7.08, $p < .001$, CI [2.60 – 7.08].

In order to have an estimation of the development of written 3SAcc over time, the sample of those child learners who produced obligatory contexts at the three testing times was taken into consideration (Comp n = 14, Collab n = 21, FFI+Collab n = 12). An analysis of this kind with regards to POSSAcc was not possible, as very few participants produced these forms repeatedly at T1, T2 and T3 (Comp n = 1; Collab n = 3; FFI+Collab n = 3). Thus, a repeated-measures ANOVA was run on 3SAcc (see Figure 24 for the interaction plot). Although no significant main effects or interactions were found, it is interesting to note that FFI+Collab had a different accuracy pattern to that of Comp and Collab. Although at T1 FFI+Collab scored the lowest 3SAcc, after the dictogloss intervention, at T2, the ratio raised to the highest (to approximately 30% mean accuracy), and finally at T3 it decreased slightly, although it was again the highest of all three experimental conditions. On the other hand, Comp, who had the highest 3SAcc at T1, lowered that rate at T2 and kept it to the same level at T3, whereas Collab experienced a decrease from T1 to T2, and a slight increase from T2 to T3. In any case, neither Comp or Collab seem to have benefitted from the DG task in terms of 3SAcc, whereas there is some reason to believe that FFI+Collab did.

Figure 24 of the repeated-measures ANOVA for 3S. Error bars show 95% confidence intervals



To conclude, as far as the analytic rubric dimensions are concerned, Table 53, Table 54 and Table 55 contain the descriptive results. For each dimension, as usual, a repeated-measures ANOVA was conducted.

Starting with the two task assessment dimensions from the rubric (Adq and Coher), it can be observed that there was a general upward trend from T1 to T2, which was clearer in the case of FFI+Collab. In the two dimensions, it was Collab that obtained the highest mean scores at the three testing times, followed by Comp and, finally, FFI+Collab. In Adq, there was a slight decline for Comp and FFI+Collab from T2 to T3, while Collab continued to increase slightly their average score. However, in Coher, all three groups experienced a decline from T2 to T3. The statistical analyses revealed some significant main effects and interactions. In Adq, there was a significant main effect of time and experimental condition, with a small effect size in both cases: $F = 4.35$, $p = .015$, $\eta^2 = 0.01$; $F = 4.76$, $p = .011$, $\eta^2 = 0.07$. The post-hoc analyses showed that there was a significant difference between T1 and T2: t = -2.88, $p = .014$, CI [-0.43 – -0.04]. Moreover, Collab scored significantly better in Adq than FFI+Collab, regardless of time: t = 3.08, $p = .009$, CI [0.10 – 0.87]. In Coher, there was also a main effect of time and experimental condition, with a small effect size in both cases: $F = 3.66$, $p = .028$, $\eta^2 = 0.02$, $F = 5.34$, $p = .007$, $\eta^2 = 0.08$. The post-hoc analysis failed to find any significant difference in the pair-wise comparisons regarding time. Yet, with regards experimental condition, the test showed that Collab scored significantly better in Coher than FFI+Collab: t = 0.37, $p = .005$, CI [0.12 – -0.87].

The language dimensions from the rubric are now analyzed. In Cohes, while Comp and FFI+Collab barely experienced any change in their mean rates across time, Collab increased slightly their rate from T1 to T2, but they decreased it again at T3. The statistical analysis failed to show any significant differences. As far as Acc is concerned, after the dictogloss intervention, the only group that increased their mean score was FFI+Collab, reflecting the same pattern we had already observed in the general accuracy text-based measures (GramErr100 and GramErrFreeC) as well as the specific dimensions (3SAcc and POSSAcc). Yet, at T3, this group's mean Acc score declined. On the other hand, in Comp and Collab, there was a slight decline in their mean score from T1 to T2, and a slight increase from T2 to T3. The statistical analysis revealed a significant interaction between time and experimental condition, with a small effect size: $F = 3.76$, $p = .006$, $\eta^2 = 0.02$. Nevertheless, the post-hoc analysis failed to indicate any significant differences.

As far as Mech is concerned, Comp and Collab progressively decreased their mean scores from T1 to T3, while the opposite was true for FFI+Collab. The statistical analysis revealed a significant interaction between time and experimental condition, with a small effect size: $F = 4.62$, $p = .002$, $\eta^2 = 0.03$. In fact, at T1, Collab scored significantly higher than FFI+Collab: t = 4.41, $p < .001$, CI [0.25 – 1.67]. Furthermore, Collab's Mech mean score was significantly higher at T1 than at T3: t = 3.41, $p = .03$, CI [0.02 – -1.02]. Finally, regarding Lex, akin to what we had previously observed in GI, there was little variation in Comp and Collab's mean scores over time, whereas in FFI+Collab there was an increase from T1 to T2. Furthermore, in line with GI values, Lex also reflected Collab's clear advantage over the rest of the conditions. In fact, the statistical analysis showed a significant main effect of experimental condition, with a small effect size: $F = 3.95$, $p = .023$, $\eta^2 = 0.07$. The post-hoc analysis indicated that, regardless of testing time, Collab obtained significantly higher Lex scores than FFI+Collab: t = 2.63, $p = .03$, CI [0.03 – 0.78].

To conclude, the analysis of child learners' individual L2 writing before and after experiencing a DG intervention did not allow us to observe great changes. Nonetheless, the comparison of T1 and T2 provided some evidence of improvement. For instance, child learners produced significantly fewer preclauses and also obtained higher adequacy scores at T2. What is more, the improvement in grammar accuracy appeared to be clearer in the case of FFI+Collab, as at T2 this group also produced significantly more full-fledged clauses. Furthermore, although not significantly, they decreased the number of grammatical errors, increased the number of error-free clauses and raised their 3SAcc. Yet, it should also be noted that most of these improvements did not occur at T3.

Apart from that, from a general view of the rubric scores, two main conclusions can be drawn. Firstly, just as in the DG writing analysis (Day 1 and Day 2, described above), the lowest mean scores in the three experimental conditions corresponded to the dimensions of cohesion and accuracy. Secondly, Collab obtained the highest mean scores at the three testing times (with the exception of Coher at T3, where Comp scored slightly higher), and the statistical analyses also confirmed their advantage over FFI+Collab (regardless of the testing time) in adequacy, coherence and lexis.

Table 53 Rubric measures at T1 / experimental group

	Comp (n = 25)			Collab (n = 25)			FFI+Collab (n = 28)		
	M	**SD**	**95% CI**	**M**	**SD**	**95% CI**	**M**	**SD**	**95% CI**
Adq	1.56	0.65	[1.29 – 1.83]	1.92	0.76	[1.61 – 2.23]	1.25	0.58	[1.02 – 1.48]
Coher	1.60	0.71	[1.31 – 1.89]	2.04	0.73	[1.74 – 2.34]	1.32	0.61	[1.08 – 1.56]
Cohes	1.36	0.70	[1.07 – 1.65]	1.64	0.76	[1.33 – 1.95]	1.29	0.60	[1.05 – 1.52]
Acc	1.52	0.77	[1.20 – 1.84]	1.76	0.66	[1.49 – 2.03]	1.21	0.42	[1.05 – 1.38]
Mech	1.96	0.89	[1.59 – 2.33]	2.28	0.84	[1.93 – 2.63]	1.32	0.67	[1.06 – 1.58]
Lex	1.52	0.65	[1.25 – 1.79]	1.88	0.78	[1.56 – 2.20]	1.32	0.61	[1.08 – 1.56]

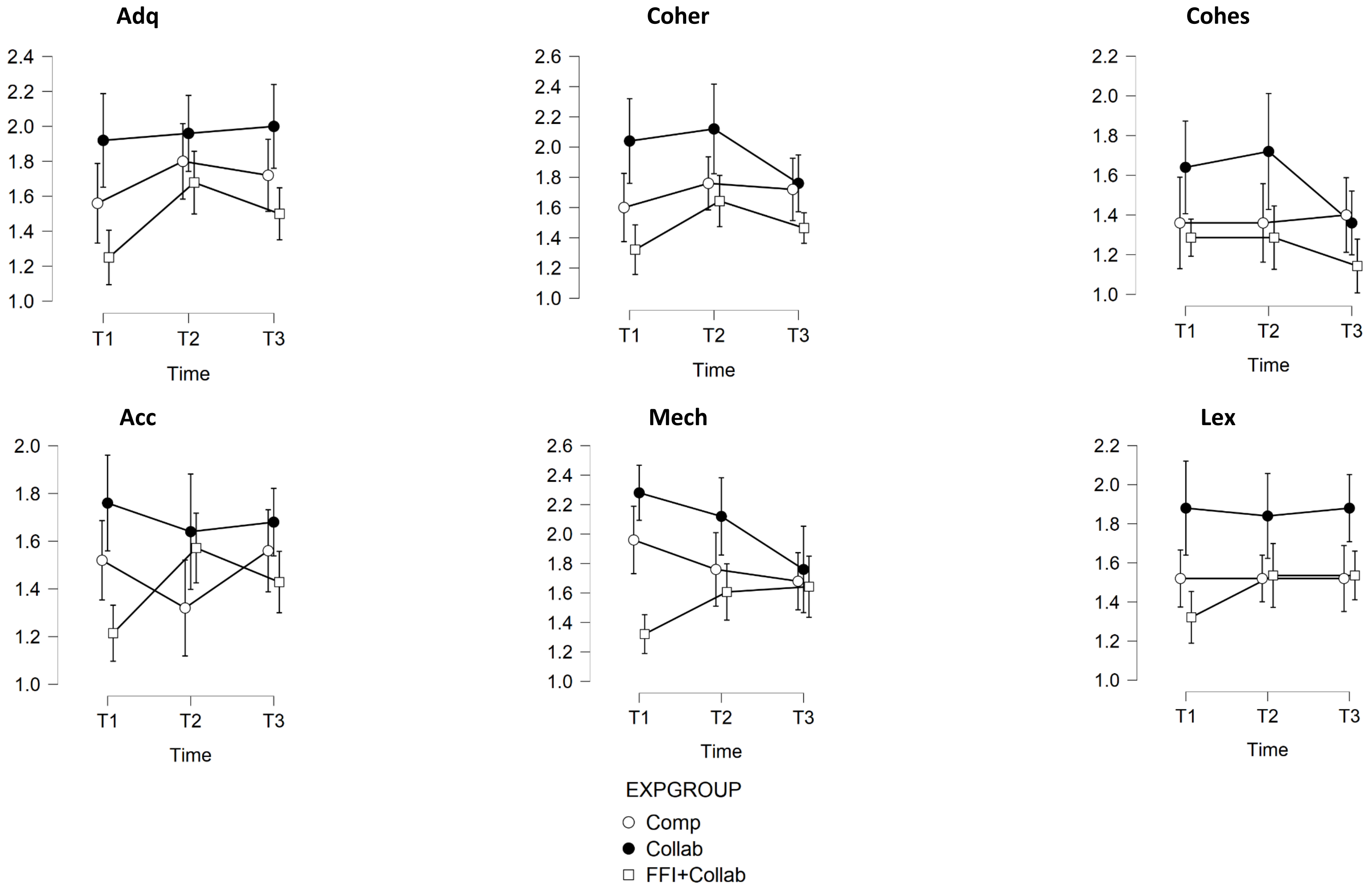
Table 54 Rubric measures at T2 / experimental group

	Comp (n = 25)			Collab (n = 25)			FFI+Collab (n = 28)		
	M	**SD**	**95% CI**	**M**	**SD**	**95% CI**	**M**	**SD**	**95% CI**
Adq	1.80	0.76	[1.48 – 2.12]	1.96	0.68	[1.68 – 2.24]	1.68	0.72	[1.40 – 1.96]
Coher	1.76	0.60	[1.51 – 2.01]	2.12	0.78	[1.80 – 2.44]	1.64	0.73	[1.36 – 1.93]
Cohes	1.36	0.57	[1.13 – 1.59]	1.72	0.79	[1.39 – 2.05]	1.29	0.66	[1.03 – 1.54]
Acc	1.32	0.56	[1.09 – 1.55]	1.64	0.64	[1.38 – 1.90]	1.57	0.74	[1.28 – 1.86]
Mech	1.76	0.72	[1.46 – 2.06]	2.12	0.83	[1.78 – 2.46]	1.61	0.83	[1.28 – 1.93]
Lex	1.52	0.65	[1.25 – 1.79]	1.84	0.62	[1.58 – 2.10]	1.54	0.64	[1.29 – 1.78]

Table 55 Rubric measures at T3 / experimental group

	Comp (n = 25)			Collab (n = 25)			FFI+Collab (n = 28)		
	M	**SD**	**95% CI**	**M**	**SD**	**95% CI**	**M**	**SD**	**95% CI**
Adq	1.72	0.79	[1.39 – 2.05]	2	0.76	[1.68 – 2.32]	1.50	0.64	[1.25 – 1.75]
Coher	1.72	0.79	[1.39 – 2.05]	1.76	0.60	[1.51 – 2.01]	1.46	0.64	[1.22 – 1.71]
Cohes	1.40	0.65	[1.13 – 1.67]	1.36	0.57	[1.13 – 1.59]	1.14	0.45	[0.97 – 1.32]
Acc	1.56	0.77	[1.24 – 1.88]	1.68	0.31	[1.45 – 1.91]	1.43	0.74	[1.14 – 1.72]
Mech	1.68	0.75	[1.37 – 1.99]	1.76	0.83	[1.42 – 2.10]	1.64	0.73	[1.36 – 1.93]
Lex	1.52	0.59	[1.28 – 1.76]	1.88	0.60	[1.63 – 2.13]	1.54	0.69	[1.27 – 1.80]

Figure 25 Interaction plots of the repeated-measures ANOVA for rubric measures. Error bars show 95% confidence intervals



6.3 RQ3 Individual differences

In this section the results from the third set of RQs will be presented. These questions revolve around child learners' affective factors (pretask and posttask attitudes), which were determined by means of different instruments. In addition, the capacity of an attitudinal variable (namely, attitudes towards writing) and L1 writing skills to predict L2 writing achievement will be explored.

a) What are YLs' attitudes to writing and collaborative work?
b) What are YLs' insights about the dictogloss task?
c) What is the impact of L1 writing skills and attitudes towards writing on L2 writing?

In order to answer RQ3a, we analyzed the attitude questionnaire (AQ) administered one week prior to the DG intervention, (see "Procedure", in section 5.4.2). Participants were allotted 30 minutes to complete it. Although the AQ was provided in the learners' L1, the analysis of the items will be presented in English for the reader's convenience. Before displaying the results, the procedure followed for the construction and validation of our instrument will be thoroughly explained. Based on previous literature (see section 4.2.1.2 above) and after conducting a piloting with twenty 5th-year-Primary learners (as explained above in "Instruments", in section 5.4.1.5), 44 five-point Likert-scale items were selected from five attitudinal areas: attitudes towards writing in the L1 (Spanish), attitudes towards writing in the L2 (English), attitudes towards pair work, attitudes towards collaborative writing and attitudes towards L2 learning.

In the current instrument, there were items used by previous questionnaires in the literature as well as others created "ad hoc" by the researcher. For example, for attitudes towards L1 and L2 writing, several items from Aula Blasco's (2016) battery of questions were taken as a reference, specifically, those targeting the subconstructs of "anxiety" and "self-efficacy". Although the original study only examined L2 writing attitudes, for our study, these questions were also adapted to the L1 writing context. The items for these two attitudinal areas were designed in parallel but they varied in their wording (positive/negative): for instance, Item 18 in "L2 writing" *"I have trouble in* *finding mistakes in my English compositions"* was comparable to Item 8 in "L1 writing" *"I am capable of finding and correcting mistakes in my Spanish compositions".*

Moreover, since Aula Blasco's (2016) participants were 17-18 years old, the language of our items was simplified, to make them more comprehensible for child learners. For example, the original question *"I often feel panic (trembling, perspiring, feeling my body rigid, having my thought jumbled, etc.) when I write English compositions under time constraint"* (i.e. long and negatively worded) was changed to *"I feel calm when I have a time limit to write in English"* (i.e. shorter and positively worded).

Finally, also bearing in mind the characteristics of our setting, we selected those questions that could be more relevant for children's learning experience. For instance, within the dimension of "L2 learning", although the original instruments (Doiz et al., 2014; Kopinska & Azkarai, 2020) contained more items targeting instrumental orientation (more specifically, two questions related to the importance of English for learner's future career and academic studies), we decided to opt only for those with which children could feel more identified (possibilities offered by English to understand films, videos and chat on the Internet).

Table 56 displays the initial forty-four items grouped by attitudinal dimension that learners were presented with, together with the citation of the reference source material. Finally, it also includes the preliminary internal consistency coefficients for the initial subdimensions.

Table 56 AQ question items

Writing in Spanish Cronbach's α = 0.713	1. I feel comfortable writing in Spanish (adapted from Aula Blasco, 2016) 2. Having a time limit when writing in Spanish makes me nervous[a] (adapted from Aula Blasco, 2016) 3. If my composition in Spanish is going to be graded, I get nervous[a] (adapted from Aula Blasco, 2016) 4. I feel concerned about what my classmates may think about my compositions in Spanish[a] (adapted from Aula Blasco, 2016) 5. I always think that I write worse in Spanish than my classmates[a] (adapted from Aula Blasco, 2016) 6. When I write in Spanish, I can easily convey my ideas without going off the point (adapted from Aula Blasco, 2016) 7. Every time I finish a composition in Spanish, I always revise it 8. I am capable of finding and correcting mistakes in my Spanish composition (adapted from Aula Blasco, 2016) 9. I devote enough time to the Spanish compositions they make me write in class 10. I can write a correctly organized text in Spanish (adapted from Aula Blasco, 2016)
Writing in English Cronbach's α = 0.756	11. Writing in English makes me nervous[a] (adapted from Aula Blasco, 2016) 12. I feel calm when I have a time limit to write in English (adapted from Aula Blasco, 2016) 13. I feel calm when my English composition is going to be graded (adapted from Aula Blasco, 2016) 14. I don't care what my classmates think about my English compositions (adapted from Aula Blasco, 2016) 15. I always think that I write worse in English than my classmates[a] (adapted from Aula Blasco, 2016) 16. When writing in English, I can always come up with ideas about the composition topic (adapted from Aula Blasco, 2016) 17. After finishing my compositions in English, I always revise them 18. I have trouble in finding mistakes in my English compositions[a] (adapted from Aula Blasco, 2016) 19. I always try to do the English compositions we have to write in class as fast as possible[a] 20. I find it hard to write a composition in English with organized ideas[a] (adapted from Aula Blasco, 2016)
Collaborative work Cronbach's α = 0.692	21. I think working in pairs is useful (adapted from Fernández Dobao & Blum, 2013) 22. I like working with my peers (adapted from Baleghizadeh & Farhesh, 2014) 23. I prefer working individually in difficult tasks[a] 24. I learn more working with my peers than on my own 25. I focus more when I am working on my own than with other peers 26. I prefer to be the one choosing who I get to work with in pairs 27. I feel comfortable working in pairs with peers than I have not chosen (adapted from Baleghizadeh & Farhesh, 2014) 28. I'd like to work more in pairs in class
Collaborative writing Cronbach's α = 0.665	29. I can come up with more ideas when writing individually than in pairs[a] 30. The compositions that we write in pairs are more accurate than the ones I write individually (adapted from Fernández Dobao & Blum, 2013) 31. I feel nervous when I have to share my ideas when writing in pairs[a] (adapted from Baleghizadeh & Farhesh, 2014) 32. I feel calmer when I have to write a text individually than in pairs[a] 33. I usually pay more attention to the text when I am writing individually than in pairs[a] 34. When we write something in pairs, I speak about things that are not related to the text[a]

	35. I learn things when writing with other peers 36. I think that the text can be better when writing in pairs than on my own (adapted from Fernández Dobao & Blum, 2013)
English learning Cronbach's α = 0.378	37. I like my English lessons at school (adapted from Doiz et al., 2014; Kopinska & Azkarai, 2020) 38. I like using English in class (Doiz et al., 2014; Kopinska & Azkarai, 2020) 39. My parents think that I should devote more time to English (Doiz et al., 2014; Kopinska & Azkarai, 2020) 40. My parents encourage me to learn English (Doiz et al., 2014; Kopinska & Azkarai, 2020) 41. I always think that other students know more English than I do[a] (adapted from Doiz et al., 2014; Kopinska & Azkarai, 2020) 42. I feel more tense and nervous in my English class than in the Spanish one[a] (adapted from Doiz et al., 2014; Kopinska & Azkarai, 2020) 43. English will help me speak with people from other countries and cultures (Doiz et al., 2014; Kopinska & Azkarai, 2020) 44. I am learning English to understand videos, music, games and chats on the Internet (Doiz et al., 2014; Kopinska & Azkarai, 2020)

[a] These items were worded negatively and, therefore, their Likert-scale values were reversed in the analysis

The complete initial exploration of the AQ items can be found in D1 (in APPENDIX D). The sample of the AQ consisted of n = 91 child learners, as two learners failed to complete it. The initial analysis showed that the number of missing answers was very low (maximum 5.5%, in items 13 and 14). Moreover, taking Kline's (2011) -3/3 benchmark for assessing skewness, we can claim that there were no severe departures from normality. Furthermore, when assessing the response distribution per question, we could observe that there was a potential ceiling effect in items 1, 22, 26, 31 and 43, where more than 40% of respondents chose the maximum score. On the other hand, we could not find a floor effect. As recommended by the literature (Elosua, 2005), we opted for discarding those items with a high ceiling attraction level.

For assessing the internal consistency, Cronbach's alpha coefficients (α) and corrected homogeneity indices (HIc) were used. Regarding the former, we can observe in Table 56 that the areas of "Writing in English" and "Writing in Spanish" ranked above 0.7, which is considered a high measure of consistency in L2 research (Dörnyei, 2007; Plonsky & Derrick, 2016). The areas of "Collaborative work" and "Collaborative writing" showed a slightly lower value, yet still acceptable. Finally, the area of "English learning" ranked last, with a low alpha value. This lower internal consistency could be due to the fact that these items addressed different constructs in the original studies, such as "Intrinsic motivation" (items 37 and 38), "Instrumental motivation" (item 44), "Interest in foreign languages" (item 43), "Parental support" (items 39, 40) or "Anxiety" (items 41, 42).

Taking a closer look at the HIc values of each item within their areas, it could be observed that the level of homogeneity of the questions varied. According to Elosua's (2005) benchmarks[18], there were some items which ranked at the level of "Insufficient" ("L2 writing": item 19; "Collaborative Work": items 26, 27; "Collaborative Writing": item 34; and "Learning English": items 39, 41). Based on these data, the questionnaire was refined and some items were deleted. In "L2 writing", "Collaborative work" and "Collaborative writing", those aforementioned items with a low HIc value were discarded. In the case of "L2 writing", the problem of item 19 may have been related to the interpretation of "as fast as possible", which according to the researcher implied a negative self-efficacy habit (as opposed to the parallel item 9 in "L1 Writing", which was positively worded). In fact, a further Spearman correlation analysis showed that it was negatively correlated with four items of that attitudinal area (see D2 in APPENDIX D for the complete bivariate correlation analysis output, in the form of heat maps).

Regarding items 26 and 27 in "Collaborative work", they asked about learners' preference for peer selection (self- vs teacher-selected), which departed from the general theme of the rest of the questions (i.e. enjoyment and perceived benefits of pair work). The Spearman correlation analysis, once again, showed negative correlations between these two items and the rest of the items. As far as item 34 in "Collaborative writing" is concerned, the question wondered about off-task behavior during CW tasks. The Spearman correlation coefficient showed that it was negatively correlated with at least two of the other subconstruct items. Once again, the negative notion of this negatively worded question may have not been correctly interpreted by child learners, who could have believed that talking about off-task topics could be something desirable.

Finally, concerning the area of "English learning", although the items that had the lowest homogeneity index were 39 and 41, theoretically it was more sensible to discard two items that belonged to the same subconstruct (in this case, learning anxiety): items 41 and 42. Moreover, it was by deleting these two items that Cronbach's alpha showed

---

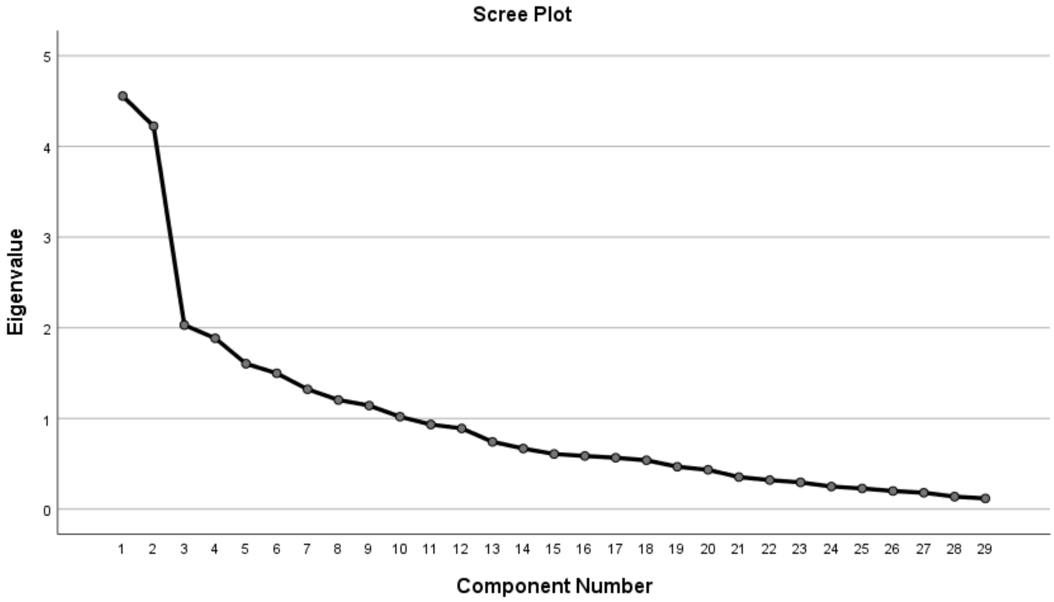
[18] Homogeneity levels (according to Elosua, 2005): $a_i < 0.19 \rightarrow$ insufficient; $0.20 < a_i < 0.29 \rightarrow$ low; $0.30 < a_i < 0.39 \rightarrow$ Good; $0.40 < a_i \rightarrow$ Very good

the largest improvement (α = 0.423), but still far below the 0.70 cut-off point which is acceptable. Therefore, a decision was made to leave “English learning” out of the subsequent analysis.

In summary, the resulting Cronbach’s alphas after the deletion of the aforementioned items were the following: L1 writing (α = 0.684), L2 writing (α = 0.776), Collaborative work (α = 0.757), and Collaborative writing (α = 0.664). The next step was to conduct a dimension reduction on the remaining 29 items of the AQ. As there was no established theoretical basis underlying the item grouping (which could have allowed a confirmatory analysis), we decided to conduct a multidimensional exploratory analysis. The method for extracting the underlying variables was a principal component analysis (PCA) (Loewen & Gönülal, 2015). The adequacy for conducting such an analysis was shown by the Kaiser-Meyer-Olkin (KMO) measure, which was .616, suggesting a moderate sampling adequacy. Moreover, Bartlett’s test of sphericity was found to be significant (p < .001), indicating that the variables were correlated, and hence, suitable for a PCA.

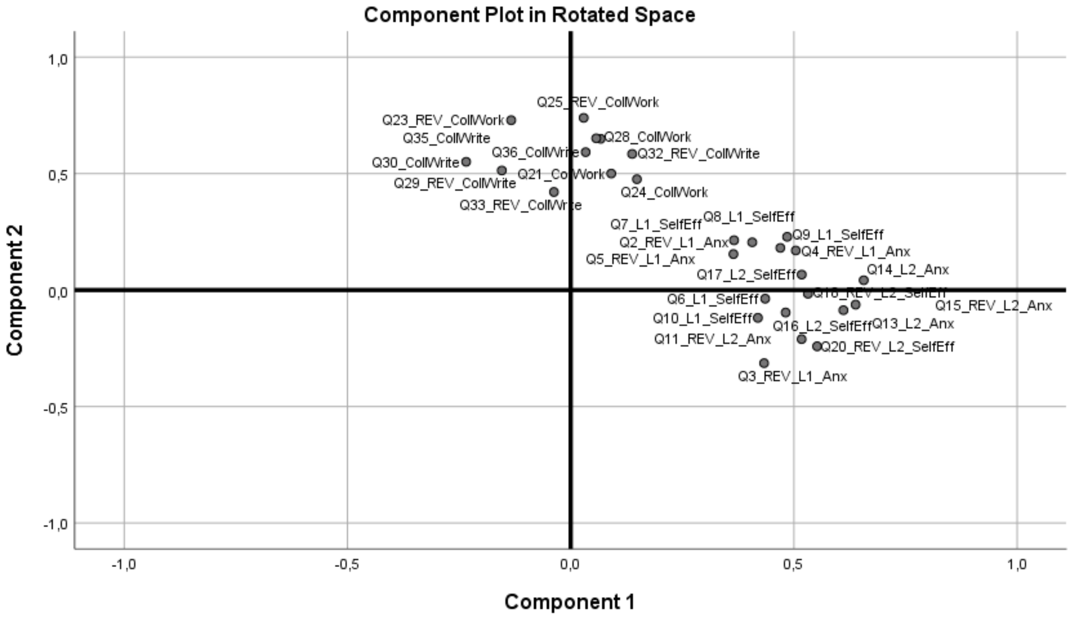
As a result of the first extraction, there were 10 underlying variables containing eigenvalues greater than 1.0, which explained 70.62% of the variance in the data (see the complete unrotated output of the first PCA round in D3, in APPENDIX D). Nevertheless, attending to the scree plot (see Figure 26), we could observe that there was a sharp drop after the second underlying component. In fact, we considered the two-factor solution the most parsimonious one, as indicated by the amount of variance explained by Factor 1 and Factor 2 (15.71% and 14.56%, respectively). Moreover, the amount of variance accounted for by the subsequent factors remained much lower (e.g. Factor 3 only added 7% more). Yet, it should be acknowledged that the cumulative amount of variance explained by our two-factor model (30.23%) was lower than the average reported in L2 research, which is close to 60% (Plonsky & Gönülal, 2015).

Figure 26 AQ PCA Scree plot



Then, we repeated the PCA with the number of factors to be extracted fixed at two and with direct oblimin factor rotation, as this method has been recommended for L2 studies (Loewen & Gönülal, 2015) and social sciences (Field, 2017). The rotated pattern matrix revealed a reliable structure of the model, as three variables of Factor 1 and four variables of Factor 2 had component loadings above .60, even so more considering the small sample size for an exploratory analysis (n = 91) and the variable-to-participant ratio in each factor (5:8) (Guadagnoli & Velicer, 1988). Moreover, no variable loaded in more than one factor and there was only one variable (item 12) which failed to load above .30 (the typical cut-off point in PCA according to Field, 2017). Therefore, the analysis was repeated leaving item 12 out. The definite model had 31.14% of cumulative amount of variance explained, that is, slightly higher than the previous model. The pattern matrix corresponding to the final model can be found in D3 (in APPENDIX D), and Figure 27 illustrates the component plot:

Figure 27 AQ component plot in rotated space



The distribution of the variables in Figure 27 allows us to see two clearly separated blocks which are associated to two well-defined factors. All the questions in Factor 1 came from the subscales of “L1 writing” and “L2 writing”, and hence, this area was defined as “Attitudes towards writing”. All items in Factor 2 came from the subscales “Collaborative work” and “Collaborative writing”, and thus, it was labelled as “Attitudes towards collaboration”. Table 57 displays the resulting subscales, loadings and extracted commonalities (h^2):

Table 57 The two subscales of AQ, question item loadings and commonalities

Subscales and question items	Component 1	Component 2	h^2
(1) Attitudes towards writing (17 items)			
Q14 I don't care what my classmates think about my English compositions	.656		.435
Q15 I always think that I write worse in English than my classmates	.638		.408
Q13 I feel calm when my English composition is going to be graded	.611		.376
Q20 I find it hard to write a composition in English with organized ideas	.552		.351
Q18 I have trouble in finding mistakes in my English compositions	.531		.282
Q17 After finishing my compositions in English, I always revise them	.517		.214
Q11 Writing in English makes me nervous	.517		.303
Q4 I feel concerned about what my classmates may think about my compositions in Spanish	.504		.290
Q9 I devote enough time to the Spanish compositions they make me write in class	.485		.297
Q16 When writing in English, I can always come up with ideas about the composition topic	.481		.237
Q5 I always think that I write worse in Spanish than my classmates	.470		.261
Q6 When I write in Spanish, I can easily convey my ideas without going off the point	.436		.190
Q3 If my composition in Spanish is going to be graded, I get nervous	.433		.275
Q10 I can write a correctly organized text in Spanish	.419		.186
Q7 Every time I finish a composition in Spanish, I always revise it	.407		.214
Q8 I am capable of finding and correcting mistakes in my Spanish composition	.366		.186
Q2 Having a time limit when writing in Spanish makes me nervous	.364		.161
(2) Attitudes towards collaboration (11 items)			
Q25 I focus more when I am working on my own than with other peers		.739	.549
Q23 I prefer working individually in difficult tasks[a]		.729	.541
Q35 I learn things when writing with other peers		.651	.431
Q28 I'd like to work more in pairs in class		.649	.430
Q36 I think that the text can be better when writing in pairs than on my own		.592	.353
Q32 I feel calmer when I have to write a text individually than in pairs		.584	.367
Q30 The compositions that we write in pairs are more accurate than the ones I write individually		.551	.347
Q29 I can come up with more ideas when writing individually than in pairs		.513	.281
Q21 I think working in pairs is useful		.500	.262
Q24 I learn more working with my peers than on my own		.476	.254
Q33 I usually pay more attention to the text when I am writing individually than in pairs		.421	.177

To conclude the validation, the internal consistency of the two subscales was checked and the following Cronbach's alpha values were obtained (see D3 in APPENDIX D for the complete output): Attitudes towards writing α = 0.838; Attitudes towards collaboration α = 0.808. Furthermore, the homogeneity of the items was also acceptable (HIc > 0.30). The whole twenty-eight question items from AQ showed an α = 0.792 and the mean item homogeneity value was .311.

Once the AQ had been validated, we calculated the total average score for each of the subscales and checked whether there were any statistically significant differences between the three experimental conditions prior to the DG intervention. After making sure that the two resulting mean scores were independent dimensions (see the Pearson correlation bivariate analysis and the dispersion plot in D4, in APPENDIX D), the descriptive statistics for each experimental condition were calculated for "Attitudes towards writing" (WritingAtt) and "Attitudes towards collaboration" (CollaborationAtt), summarized in Table 58. The *n* in each condition comes from the number of participants who did not have any missing values in the questions from WritingAtt and CollaborationAtt.

Table 58 AQ descriptive statistics / experimental group

	WritingAtt			**CollaborationAtt**		
	M	**SD**	**95% CI**	**M**	**SD**	**95% CI**
Comp (n = 18)	3.65	0.46	[3.42 – 3.88]	3.14	0.46	[2.91 – 3.37]
Collab (n = 22)	3.67	0.41	[3.48 – 3.85]	3.44	0.54	[3.20 – 3.69]
FFI+Collab (n = 26)	3.18	0.61	[2.93 – 3.42]	3.13	0.64	[2.87 – 3.39]

Next, a one-way ANOVA was run in order to check for significant differences in the subdimension scores (see the complete statistical output in D5 in APPENDIX D). Regarding WritingAtt, the test revealed a significant main effect of experimental condition, with a medium effect size: $F = 6.59$, $p = .002$, $\eta = 0.16$. In fact, learners in Comp and Collab had a significantly higher mean attitude score towards writing than their FFI+Collab counterparts: $t = 3.10$, $p = .008$, CI [0.10 – 0.81].

Conversely, the test failed to show any significant differences in the case of CollaborationAtt. Additionally, the AQ contained four extra questions in which learners could self-report the time dedicated to individual and collaborative work practices. The results, which are summarized in Figure 28, Figure 29, Figure 30 and Figure 31 below, indicated some differences depending on the school they were at.

When asked about individual and collaborative work in general (Figure 28 and Figure 29), learners in Comp and Collab, who were recruited from School A, claimed to work more often individually than in pairs, whereas the opposite was true for learners in FFI+Collab, who came from School B. When asked about their usual writing mode

(Figure 30 and Figure 31), a large number of learners in Comp and Collab stated that they wrote more often individually than in pairs. In contrast, in the case of children in FFI+Collab, although they claimed to write more individually, their proportion of time dedicated to collaborative writing came very close.

Figure 28 Self-reported time dedicated to individual work / experimental group

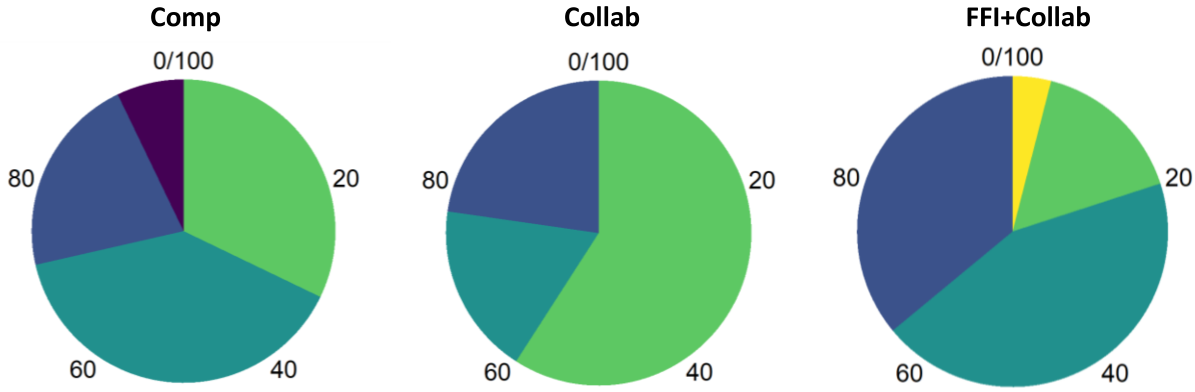


Figure 29 Self-reported time dedicated to collaborative work / experimental group

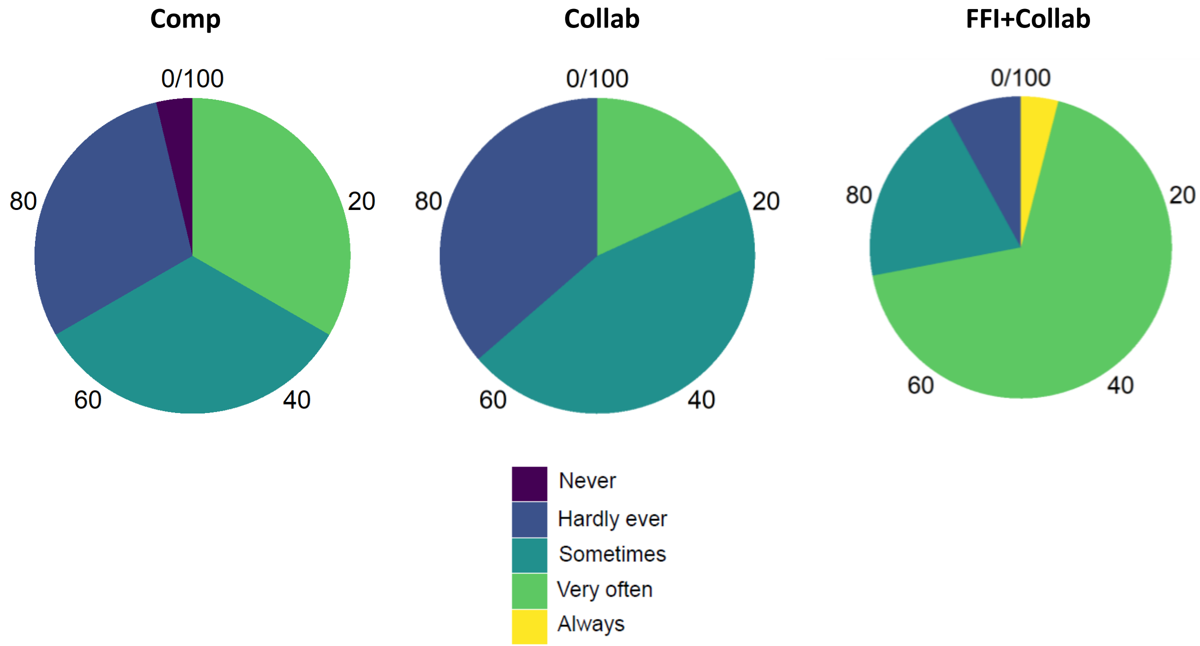


Figure 30 Self-reported time dedicated to individual writing / experimental group

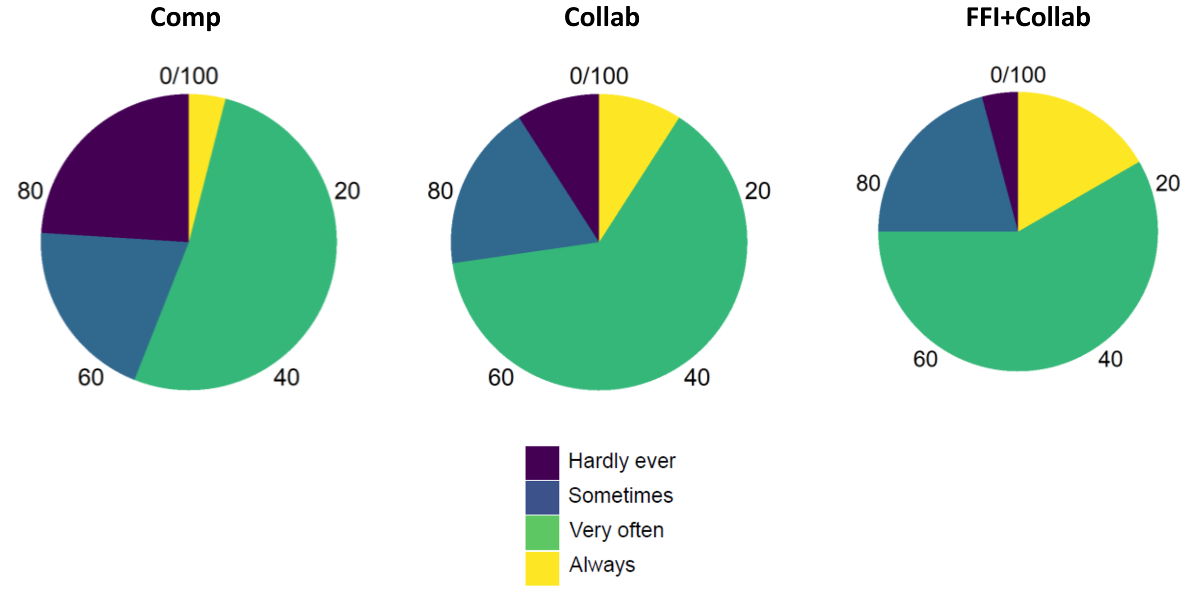
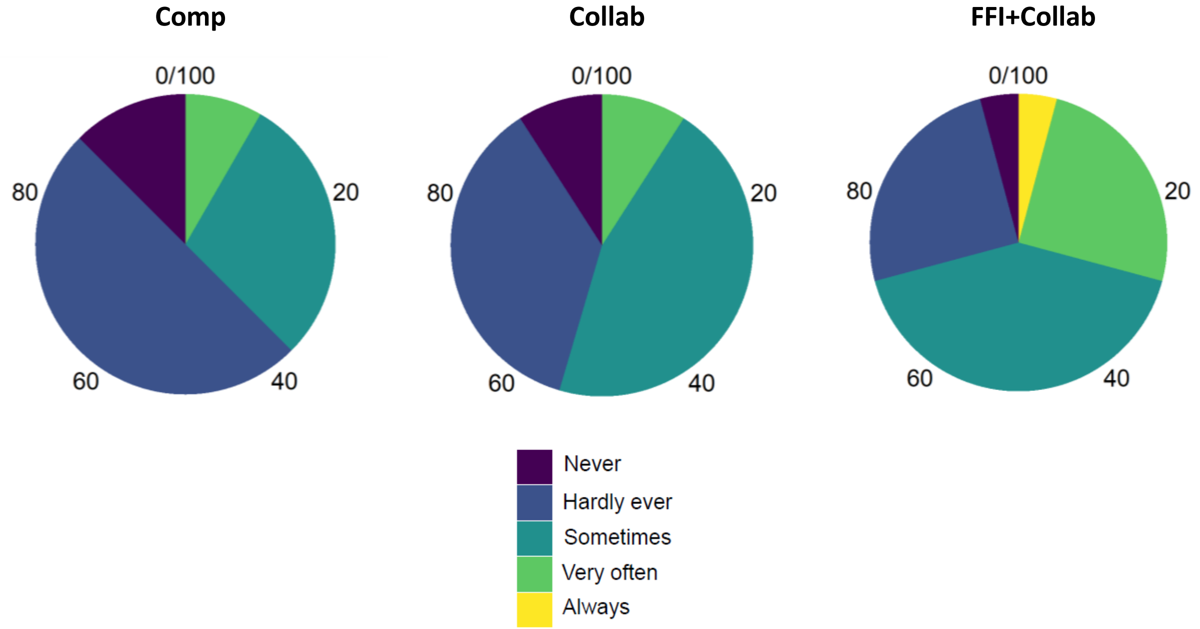


Figure 31 Self-reported time dedicated to collaborative writing / experimental group



To finish RQ3a, children's open-ended responses were analyzed, which, following Kopinska and Azkarai (2020), were designed to gather qualitative information, in our case, on children's perceptions about writing in Spanish and in English, and also about individual work and pair work. In these questions, learners had to write one word to express how the felt with regards to each of these four dimensions. These words were then classified into sentiments on NVivo (QSR International, 2012), in terms of their connotation: positive, negative, neutral and off-topic response (i.e. not related to the question). After each word children could also briefly provide the reason for their choice. In the current analysis, this justification was only used to determine the word connotation whenever it was difficult to derive it from the word per se. For instance, a word like "concentration" could reflect either a positive sentiment (e.g. when referred to individual work, since a child could feel more focused on the task in this condition) or negative sentiment (e.g. when referring to pair work, as peers could lead to off-task behavior).

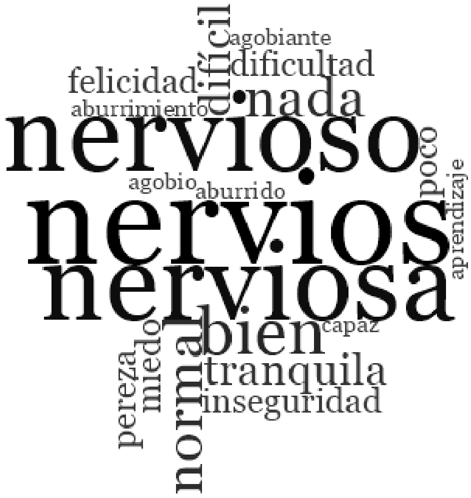
Table 59 contains the sentiment percentages in each experimental condition for L1 writing and L2 writing, while Table 60 contains the sentiment percentages for individual and collaborative work in each condition. In addition, the word clouds for the twenty most frequent words in each of the questions are provided in Figure 32 and Figure 33:

Table 59 Sentiment percentages regarding L1 and L2 writing / experimental group

	L1 writing			**L2 writing**		
	Comp	**Collab**	**FFI+Collab**	**Comp**	**Collab**	**FFI+Collab**
Positive	75	78.6	70.9	15.6	25	16.1
Negative	12.5	14.3	19.3	62.5	60.7	71
Neutral	0	7.1	9.7	9.4	14.3	12.9
Off-topic	12.5	0	0	12.5	0	0

Figure 32 Word clouds for L1 and L2 writing

L1 writing



L2 writing

As can be observed in Table 59, regardless of the experimental condition, child learners' open-ended responses reflected primarily a positive sentiment in relation to L1 writing (above 70-80%), whereas more than half of their answers regarding L2 writing expressed a negative feeling (60-70%). This clear predominance of one of the two sentiments in L1 and L2 writing is also corroborated by the word cloud (Figure 32). Regarding L1 writing words, among the most repeated ones appear *"bien"* (good), *"tranquilidad"* (calm), *"feliz"* (happy), "*alegría"* (joy) or *"facilidad"* (ease) and *"comodidad"* (comfort). As far as L2 writing is concerned, the most repeated word is *"nervios"* (nerves, anxiety) and its derivates *"nervisoso/a"* (nervous). We can also observe some other frequent negative words, such as *"difícil"* (difficult)*, "agobiante"* (stressful) or *"aburrimiento"* (boredom), together with a few positively connotated ones, such as *"bien"* (good) or *"capaz"* (capable).

Table 60 Sentiment percentages regarding individual and collaborative work / experimental group

	Individual work			**Collaborative work**		
	Comp	**Collab**	**FFI+Collab**	**Comp**	**Collab**	**FFI+Collab**
Positive	53.1	64.3	51.6	68.7	82.1	74.2
Negative	37.5	21.4	22.6	12.5	10.7	16.1
Neutral	3.12	10.7	22.6	15.6	3.6	6.4
Off-topic	6.2	3.6	3.2	3.1	3.6	3.2

Figure 33 Word clouds for individual and collaborative work

Individual work Collaborative work

Regarding learners' open-ended responses to individual work, the higher percentage corresponds to positive sentiments in all three conditions, accounting for 50-65% of all answers. Yet, the negative and neutral responses also represent a large proportion (ranging 30-45%). Such dichotomy in children's answers can also be observed in their most frequent words (Figure 33), as *"bien"* (good), *"tranquilidad"* (calm) or *"concentración"* (concentration) appear next to words such as *"normal"* (normal), *"aburrimiento"* (boredom) and *"inseguro"* (insecure). In contrast, regarding their responses to collaborative work, a positive sentiment predominates, whilst negative and neutral responses account for less than 30%. Thus, the most frequent words were *"bien"* (good), *"contento/a"* (happy), *"alegría"* (joy), *"a gusto*" (comfortable) or *"tranquilo"* (calm).

There were not large differences in children's responses with regards to their experimental condition prior to the DG intervention. Yet, we can observe that, just as in the case of the scores from the two Likert-scale dimensions in AQ (explained above), learners in Collab tended to reflect more positive sentiments than learners in Comp and FFI+Collab.

RQ3b will now be considered, that is, the analysis of child learners' perceptions of the DG task. The dictogloss questionnaire (DQ) was completed by the participants one week after they carried out the second DG task, on the same day as the writing posttest (T2, see above). Children had approximately 30 minutes to fill out the DQ in their L1. As in the case of AQ, first the validation of the questionnaire will be explained and, afterwards, any differences between the three experimental conditions will be analyzed (Comp, Collab and FFI+Collab).

DQ contained questions that were common to all the experimental conditions, and others that were specific of each condition. The items were inspired in questions from the studies by Shak (2006) and Shak and Gardner (2008), which targeted young ESL learners' perceptions of focus on form tasks, including DG. Yet, we also introduced new items that addressed more specifically different aspects related to the task in the current study, such as children's perception of task repetition (TR) or task setting (individual vs collaborative). Table 61 shows the complete battery of questions that our participants were provided with, divided into two blocks: items 1-7, the common block (i.e. questions that all children answered) and items 8-16, the specific block (i.e. questions that were related to each experimental condition).

Table 61 DQ question items

Common (α = 0.585)	1.	I found the tasks too difficult[a]
	2.	I enjoyed the listening part
	3.	I enjoyed the part where I had to take down notes while I listened to the audio
	4.	I enjoyed the rewriting part
	5.	I think that I did the task better on the second day than on the first
	6.	The more I do the task, the better I'll be at it
	7.	I felt more motivated to do the task on the second day than on the first
Individual setting (Comp) (α = 0.323)	8.	I'd like to do the tasks again in the normal lessons with my teacher just as I did them now
	9.	I'd like to do the tasks individually again
	10.	I felt at ease working on my own
	11.	I enjoyed the part where I had to write down my doubts and thoughts on a photocopy
Collaborative setting (Collab and FFI+Collab) (α = 0.742)	12.	I'd like to do the tasks again in the normal lessons with my teacher just as I did them now
	13.	If we do the tasks again, I'd like to be in pairs again
	14.	I felt at ease working with my peer
	15.	I enjoyed the part where I had to speak to my peer
Only FFI	16.	I enjoyed the grammar exercises before the writing task

[a] Negatively worded item

The complete initial exploration of the DQ question items can be found in D6 (in APPENDIX D). The learner sample of DQ consists of 84 children from the original 93. In fact, there were two missing participants the day the DQ was administered, and seven participants' responses were discarded, as they had failed to carry out the DG task on one of the two experimental days (and hence, their answers could be somewhat biased). In this exploration, it may be observed that in general the number of missing responses was low (below 8%) and, by looking at the skewness values, it may be claimed that there were not strong deviations from normality (well within the -3/3 range, Kline, 2011).

With regards to the common block, Cronbach's alpha for internal consistency was below the desired benchmark (< 0.70). A closer look allowed us to determine that there were two problematic items. First, the negatively worded item 1 correlated negatively with the rest of the items of the block and showed a very low homogeneity index (HIc < 0.19). Secondly, item 6 had a ceiling effect, as 41.7% of the participants chose the highest score, while 1.2% chose the lowest. What is more, the HIc was also equally insufficient. In fact, we could argue that the content of item 1 was more related to the construct of perceived task difficulty than to the concept of general task preference (Wang & Li, 2019), and we decided to analyze it independently. Regarding item 6, although it dealt with the concept of repetition, the fact that it did not ask about learners' task preference but about the future influence of the task (Wang & Li, 2019) could account for the low correlation with items 5 and 7 (see the Spearman correlation heatmap for each subscale in D7, in APPENDIX D). After removing item 1 and item 6, Cronbach's alpha was improved to 0.708, which we considered acceptable.

As far as the items targeting the individual setting are concerned (which only participants in Comp completed, n = 26), Cronbach's alpha ranked again well below the adequate threshold. In this case, item 11 was identified as the problematic question, since it showed a negative and very low HIc with the rest of the subscale items. Moreover, the content of item 11 is somewhat detached from the rest of the items in this block, as it does not enquire directly about learners' feelings towards individual task work, but rather to one of the task parts (i.e. the written languaging worksheet). Therefore, we decided to analyze item 11 separately, akin to item 16, which asked about the pretask FFI. After removing item 11, Cronbach's alpha for the Individual subscale raised to 0.501. Yet, it must be acknowledged that this value was somewhat below the sufficient benchmark.

Finally, regarding the collaborative setting items (which participants from Collab and FFI+Collab completed, n = 58), Cronbach's alpha value was above 0.70, and all four items had a good level of HIc (> 0.30, Elosua, 2005). Hence, the final version of the DQ consisted of 13 items divided into three main blocks (Common: 5 items; Individual setting: 3 items; Collaborative settings: 4 items) and two independent questions which enquired about condition-specific task features, namely, the written languaging worksheet (item 11) and the pretask FFI (item 16).

The final step for the questionnaire validation involved performing a PCA on the "Common" block items. Although the current sample size (n = 84) was below the usual minimum sample recommendation (n = 100) for this sort of analyses (Plonsky & Gönülal, 2015), the KMO value was .662 and Bartlett's test of sphericity was found to be significant ($p < .001$), suggesting that a PCA was possible on the data. Due to the smaller sample size in the "Individual setting" block (n = 26) and the "Collaborative setting block" (n = 58), it was not possible to conduct a PCA and we only took into consideration the aforementioned internal consistency Cronbach alpha values.

A unidimensional analysis on the common block items was run, as all items in the Common subscale were expected to measure the same underlying factor (i.e. DG task preferences). Bearing in mind the characteristics of the data (Loewen & Gönülal, 2015), as in the case of AQ, we opted for an oblique direct oblimin rotation. The complete statistical output can be found in D7, in APPENDIX D. The results of the component extraction indicated that, contrary to our initial assumption, there were two main variables that explained 71.88% of the variance in the data (46.60% and 25.27%, respectively). Furthermore, the scree plot (Figure 34) also showed a progressive decrease from Factor 1 to Factor 2, while after that point the Eigenvalues were below 1. The component plot (Figure 35) also indicated that there were two independent blocks, which are to a large extent justifiable by theory: Factor 1 deals with DG task general preferences, while Factor 2 gauges learners' perceptions towards task repetition. The factor loadings of each item can be found in Table 62:

Figure 34 DQ PCA scree plot

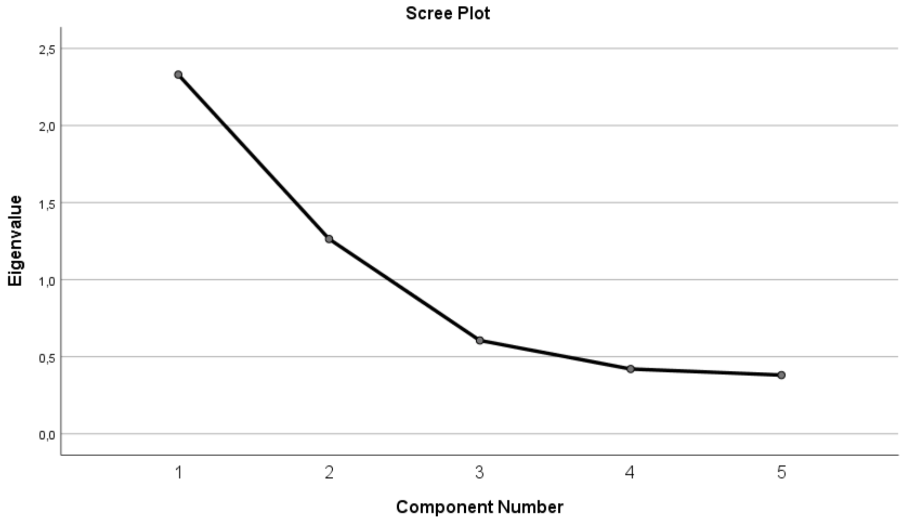


Figure 35 DQ component plot in rotated space

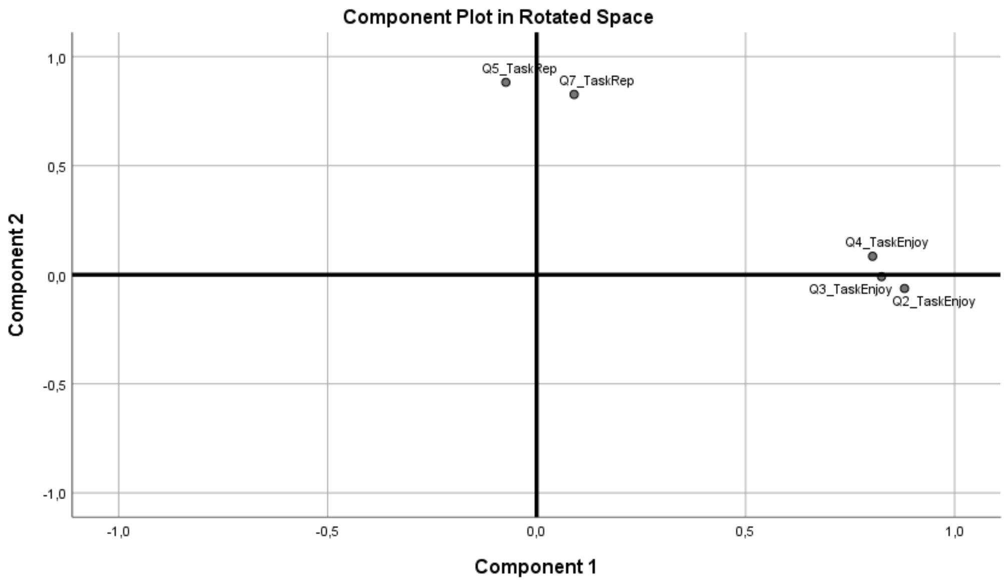


Table 62 The two subscales of DQ, question item loadings and commonalities

Subscales and question items	Component 1	Component 2	h^2
(1) DG task preferences (3 items)			
Q4 I enjoyed the rewriting part	.801		.753
Q2 I enjoyed the listening part	.790		.678
Q3 I enjoyed the part where I had to take down notes while I listened to the audio	.708		.685
(2) Perception of TR (2 items)			
Q5 I think that I did the task better on the second day than on the first		.759	.753
Q7 I felt more motivated to do the task on the second day than on the first		.656	.726

The internal consistency of the two subscales and the following Cronbach's alpha values were obtained (see "Final internal scale validation" in D7, in APPENDIX D). The results show a very good internal consistency for the "DG task preferences" subscale ($\alpha = .789$) and a slightly lower value for "Perceptions of TR" ($\alpha = .637$). The homogeneity index values in each subscale were also very good or higher (HIc > 0.40) (Elosua, 2005).

Next, the total subdimension mean scores for each block were calculated and it was checked to see whether they were independent from each other by means of Pearson correlation and dispersion plots. According to Plonsky and Oswald (2014), the correlation coefficients were low (close to .25 or below), except for the relationship between DG preferences and perceptions of collaborative setting, which ranked medium (close to .40). To conclude, we calculated the descriptive statistics for each experimental condition and checked whether there were any statistically significant differences between them. The sample in each experimental condition comes from the number of participants who did not have any missing values when calculating the final subdimension scores. A series of one-way ANOVAs were conducted in each subdimension to determine whether the differences in the average scores were statistically significant (see the complete ANOVA output in D10, in APPENDIX D)

As can be observed in Table 63, regarding the first attitudinal dimension ("Task preferences"), learners in Collab and FFI+Collab displayed higher mean scores than their peers in Comp. In fact, the one-way ANOVA analysis found a significant main effect of experimental condition, with a medium effect size: $F = 14.22$, $p < .001$, $\eta = 0.27$. The post-hoc analysis confirmed that Collab had a higher mean attitude towards the DG task features than Comp and FFI+Collab: Collab vs Comp $t = 5.27$, $p < .001$, CI [0.70 – 1.87]; Collab vs FFI+Collab $t = 3.40$, $p = .003$, CI [0.24 – 1.35]. With regards to the perception of TR and task difficulty, all learners regardless of their experimental condition displayed similar average scores, although Collab was again the group that obtained the highest mean value. The analysis failed to show any significant differences.

As far as the two working modes are concerned, the individual setting obtained a lower mean score than the collaborative setting, and within the latter, there were also some differences between the two experimental conditions (Collab vs FFI+Collab). Moreover, experimental condition proved to be statistically significant, with a medium effect size: $F = 16.58$, $p < .001$, $\eta = 0.24$. The post-hoc analysis showed that learners in Collab displayed a better perception of their own setting than those in FFI+Collab: $t = 4.07$, $p < .001$, CI [0.37 – 1.08].

Finally, regarding the items that enquired about two particular dictogloss features (Q11, written languaging, and Q16, FFI), learners in Comp and FFI+Collab, respectively, displayed a positive perception (higher than 3.5).

Table 63 DQ mean scores / experimental group

	Comp (n = 25)			Collab (n = 26)			FFI+Collab (n = 30)		
	M	**SD**	**95% CI**	**M**	**SD**	**95% CI**	**M**	**SD**	**95% CI**
Task preferences	2.51	0.99	[2.10 – 2.91]	3.79	0.60	[3.55 – 4.03]	3	0.96	[2.64 – 3.36]
Task repetition	3.48	1.10	[3.02 – 3.93]	3.69	0.86	[3.34 – 4.04]	3.43	1.02	[3.05 – 3.81]
Task difficulty	3.04	1.21	[2.55 – 3.53]	3.33	1.21	[2.86 – 3.81]	3.06	1.12	[2.65 – 3.48]
Individual setting	2.86	0.81	[2.52 – 3.20]	-	-	-	-	-	-
Collaborative setting	-	-	-	4.39	0.62	[4.13 – 4.65]	3.67	0.68	[3.41 – 3.92]
Written lang.	3.73	1.12	[3.28 – 4.18]	-	-	-	-	-	-
FFI	-	-	-	-	-	-	3.43	0.93	[3.08 – 3.78]

Note: Written lang. → Written languaging worksheet

These quantitative data about learners' task perceptions were supplemented by some qualitative information obtained via two open-ended questions which were included at the end of the DQ. These questions were adapted from Kopinska and Azkarai (2020) and enquired child participants to write down something they liked and something they disliked about the DG task. However, instead of requiring learners to write up to three positive and negative aspects, we decided to ask them to write down a single word in each category, since the authors in the original study had reported that children struggled to come up with that many words.

The procedure we followed to analyze these data was the same as in Kopinska and Azkarai's (2020), which was based on Garrett and Gallego Balsà (2014). First, we identified the discrete ideas in children's responses, and secondly, we classified them into broader themes in relation to the DG task. The analysis was carried out on NVivo (QSR International, 2012). As in the case of the AQ, children's word choice explanation was only used to guide our classification. Whenever child participants mentioned aspects which were not part of the DG task (e.g. the attitude questionnaire, the story continuation writing tests), these words were not taken into account for the analysis.

Likewise, words that were too vague or generic (e.g. *"todo"* [everything], *"nada"* [nothing]) were not categorized in any topic. Table 64 and Table 65 summarize the topics that the child learners from the different experimental conditions mentioned when referring to aspects they liked and disliked, respectively. The topics have been ranked depending on their frequency of appearance within each group.

Table 64 What YLs liked about the DG task / experimental group

Comp (n = 25)		Collab (n =26)		FFI+Collab (n = 30)	
Topic	***N***	**Topic**	***N***	**Topic**	***N***
+Listening/Audio recordings	6	+Pair work/Collaboration	10	+Pair work/Collaboration	8
+Learning English	2	+Listening/Audio recordings	5	+Comfort	3
+Note-taking	2	+Fun	2	+Fun	3
+Originality/Break from routine	2	+Comfort	1	+Writing/Text reconstruction	3
+Comfort	1	+Note-taking	1	+FFI	1
+Stories	1	+Stories	1	+Test	1
+Task time	1	+Writing/Text reconstruction	1		
+Test form	1				
+Writing/Text reconstruction	1				

As we can see in Table 64, when considering what aspect children liked from the DG task, learners who had carried it out in pairs (Collab and FFI+Collab) most frequently referred to collaboration and pair work. Conversely, those learners in Comp mentioned the listening stage and the audio recordings as most preferred. It is interesting to note that none of the child learners in Comp referred to individual work as such, but two participants did mention the fact that they had learnt English through the tasks, while another learner claimed to feel comfortable while performing the tasks. The listening part also ranked as one the preferred aspects for Collab, but it did not arise among learners in FFI+Collab.

Furthermore, while certain learners in the individual condition considered the tasks a novelty and a break from the routine, it was some learners in Collab and FFI+Collab that described it as fun. In addition, one learner from Comp and another from FFI+Collab reported that they enjoyed the fact that the tasks were carried out as language tests (including the recording of their performance). As we can observe, while writing and the text reconstruction was not generally mentioned as an enjoyable part, more learners in FFI+Collab mentioned this topic than in the other two groups. Finally, the FFI pretask activities were only mentioned by one learner in this group. In fact, he claimed that he had enjoyed the "-*s* activity" (MAR027), because it was like "solving a mystery".

Table 65 What YLs disliked about the DG task / experimental group

Comp (n = 25)		Collab (n =26)		FFI+Collab (n = 30)	
Topic	***N***	**Topic**	***N***	**Topic**	***N***
-Writing/Text reconstruction	6	-Writing/Text reconstruction	6	-Pair work/Collaboration	5
-Listening/Audio recordings	3	-Listening/Audio recordings	2	-Writing/Text reconstruction	4
-Difficult/Effortful	3	-Note-taking	2	-Difficult/Effortful	4
-Lack of choice/decision	2	-Task time	2	-Listening/Audio recordings	2
-Task repetition	1	-Difficult/Effortful	2	-Test	2
-Task time	1	-Lack of choice/decision	1	-Lack of choice/decision	1
-Test	1	-Pair work/Collaboration		-Task time	1
		-Test	1		

Those aspects which learners disliked about the DG task are summarized in Table 65. Learners in Comp and Collab agreed on the fact that writing and the text reconstruction stage were their least preferred parts. On the contrary, learners in Comp again mostly referred to pair work and collaboration. In fact, it should be recalled that it was participants in this group that displayed fewer collaborative patterns (see RQ2a). In their justification, children for instance argued that "it was difficult to reach an agreement" [MAR003], that there were ideas their peer disliked [MAR021] or that their peer did all the work [MAR034].

The listening and the audio recordings, which had also been mentioned as a positive aspect above, were referred to as a negative aspect by some children from the three conditions. Another common response was related to task difficulty, as some children considered the tasks too demanding. Another theme that members from all three groups pointed out was the lack of decision or choice. This complaint was mostly related to the fact that they could not choose their own partner, or in the case of learners in Comp, that they could not decide to carry out the DG task in collaboration. Finally, a few learners from the three conditions also referred to the excessive time pressure and the fact that they perceived our intervention as an exam.

In addition to the DQ, learners' perceptions about the DG task were also tapped by means of focus group interviews, which were carried out during the last week of the experimental procedure. Child participants could take part in these short sessions (10 minutes) on a voluntary basis during one of the class breaks. These interviews were semi-structured by the following three questions:

1. What do you think the goal of these tasks was? Have they been useful?
2. How did you find the tasks? How did you feel about them? Did you enjoy working individually or in pairs?
3. What do you think of the whole experimental procedure?

There were four focus groups. Two were from School A, mixed with participants from Comp (n = 9) and Collab (n = 6), and another two were from School B, exclusively consisting of participants from FFI+Collab (n = 10). The maximum number of participants in these focus groups was six and the minimum four, in order to guarantee a smooth flow of the conversation. In what follows, we will provide excerpts from the interviews to complement the data obtained from the DQ. A complete transcript of a focus group interview can be found in D11 (in APPENDIX D). The transcripts were analyzed using the software NVivo (QSR International, 2012). We coded similar topics and comments that child participants raised in response to the questions above.

With regards to the first question, learners from Comp perceived that the goal of the DG task was to improve their listening comprehension skills, whereas Collab and FFI+Collab referred to becoming better at pair work. This dichotomy is clearly reflected in excerpt (36), where ABE043 and ABE057 belong to Comp, and ABE039 to Collab:

(36)

*ABE043:	*yo creo que es para mejorar el inglés* (I think that it's for improving our English).
*RES:	*mejorar el inglés, muy bien* (improving your English, very good).
*ABE057:	*sí, ¿pues mejorar el entendimiento? no sé cómo se dice* (yeah, and improving our understanding?)
*RES:	*sí, muy bien la comprension oral, ulermena* (yes, very good, your oral comprehension).
*ABE057:	*bai.* (yes)
*RES:	*¿algo más?* (something else?).
*ABE039:	*trabajo en equipo* (team work).
*ABE057	*pero eso algunos* (but that only some of you).

In (36), learners from the individual setting consider that one of the task goals was to improve their English and, more specifically, their oral comprehension skills.

However, their peer from the collaborative setting believes that the objective was to develop their team work skills, which ABE057, from Comp, considers does not apply to all participants. Furthermore, the trend that learners in the individual condition showed already in their open-ended answers in the DQ, where they emphasized learning and listening as their preferred task aspects, was also reflected in their focus group suggestions. In fact, besides listening comprehension, learners from Comp mentioned a large array of metacognitive goals that they considered the DG targeted: *"ser imaginativos"* (being imaginative) [ABE057], *"me ha hecho entender el nivel de inglés que tengo"* (it has made me realize about my level of English) [ABE043], *"memorizar"* (memorize), *"hacernos trabajar mejor"* (make us work better) [ABE027].

Yet, although participants from Collab generally agreed that the main objective was to develop their team work skills, those children in FFI+Collab, probably influenced by the pretask grammar-focused activities, also highlighted certain linguistic goals, as can be observed in the example below:

(37)

*MAR012	*para ver qué nivel tenemos de… para ver qué nivel de inglés tenemos* (to see what our level is… our English level).
*RES	*sí, vale muy bien* (yes, very good).
*MAR014	*para aprender a trabajar en parejas, para ver cómo trabajamos en parejas en inglés* (to learn how to work in pairs, to see how we work in pairs in English).
*RES	*perfecto muy bien, ¿MAR016?* (perfect, very good)
*MAR016	*¿para ver cuánta información recogemos?* (to see how much information we can gather)
*RES	*vale, también puede ser, ¿o sea cosas de memoria? ¿cuánto recordáis?* (ok, it could be, you mean memory? how much you remember?)
*MAR012	*sí* (yes).
*RES	*vale, ¿algo más?* (ok, something else?)
*MAR001	*cuánto sabemos sobre inglés* (how much English we know).
*RES	*cuánto sabéis sobre inglés…* (how much you English you know…)
*MAR018	*para practicar* (to practice).
*RES	*¿para practicar y exactamente qué? ¿que podría ser?* (to practice what exactly? what could it be?)
*MAR018	*lo de…* (that of…)
*MAR009	*ulermena* (comprehension).
*RES	*ulermena* (comprehension).
*MAR012	*gramática* (grammar).

*MAR001 *las frases… o sea pasado presente futuro.* (the sentences… I mean, past, present or future)
*MAR009 *signos de puntuacion* (punctuation marks).

In (37), a participant from FFI+Collab mentions that the goal of the DG task was to check their level of English. Nonetheless, by engaging in a brainstorming, the children touch upon different aspects (e.g., pair work and memory training), until they narrow down the scope of the task to linguistic issues, including more formal topics such as grammar and mechanics.

With regards to the second part of the first question, that is, when child participants were asked about their perceived usefulness of the DG task, there was not a unified response even within members of the same experimental condition. For instance, these FFI+Collab participants mentioned a large number of dimensions they considered DG useful for: *"en el vocabulario, palabras nuevas"* (for vocabulary, for [learning] new words) [MAR032], *"en el listening"* (for listening) [MAR029], *"en escribir"* (for writing) [MAR024]. Regarding writing in particular, MAR012, who had a high opinion of her own proficiency, indicated the following:

(38)

*MAR012: *yo salvo la mayoría de las cosas ya las había hecho en la academia, pero los* writings *se me dan bastante mal y pues para practicarlos me ha venido bien.*
(I for one had done most of the things at the language school, but I'm quite bad at writing, and in that sense it has come in handy)
*RES: *¿alguien más? ¿A alguien se le ocurre…?* (anybody else? Anybody can think of…?)
[…]
*MAR012 *¡ah, sí! Que ha venido bien que vengas porque el inglés en el colegio es muy bajo, o no sé si nos dan las clases porque creen que es nuestro nivel o porque es lo que saben dar pero me parece que es muy bajo…*
(oh, yeah! It's been good that you came because the level of English at our school is very low, I don't know if they teach that kind of lessons because they believe that's our level or because they can't teach in another way, but I consider it very low…)
*RES: *tú te sientes que el nivel que exigen…* (so you reckon that the level they ask for…)
*MAR012 *es para niños de cinco años y a mucha gente le parece eso y no sé por qué no nos dan algo más, que somos niños de sexto, el inglés es muy bajo y*

esto a mucha gente le ha pillado de sopetón, porque es un poco más alto que este (,) no es superalto pero es más alto que este.
(it's for five-year-old children, and many people think so too, I don't understand why they don't give something more [challenging], because we are 6th year children, and the English level is very low. That's why many people have been taken by surprise, because the level [of the experiment] is a little higher than ours, it's not superhigh, but it's definitely higher than ours)

In (38), MAR012 claims that, while the tasks did not constitute something new for her (as she attended extracurricular lessons), she appreciated the opportunities for practicing writing, the domain she finds most difficult. Moreover, when elaborating her contribution to the discussion, this participant describes the experimental intervention as useful as they could do something that suited their level of English better, and adds that the activities they usually do at school are too easy.

This opinion is linked to the second question, which enquired child learners about their attitudes towards the task, pair work and individual work. The idea of the tasks being difficult was raised more often by participants from FFI+Collab than those in Comp and Collab. However, learners from Comp and FFI+Collab expressed that the audio recording went too fast, as can be observed in their comments in (39):

(39)

(39a) Comp

*ABE009:	*bien, muy bien* (good, very good).
*ABE027:	*a mí no me han gustado mucho porque era de escuchar audios* (I didn't like [the tasks] very much because it involved listening to audios).
*RES:	*¿era escuchar audios y no te gusta el listening?* (so it included listening to some recordings and you don't like that part)
*ABE027:	*no* (no).
*RES:	*no te gusta, vale (,) ¿ABE001?* (you don't like that, ok, ABE001?)
*ABE001:	*ni bien ni mal, un término medio* (so so, average).
*RES:	*vale, bien* (ok, good).
*ABE015:	*yo igual que* ABE027*, también que no me ha gustado mucho* (I feel the same as ABE027, I didn't like them very much).
*RES:	*¿por el tema de escuchar o por otra cosa?* (because of the audio recordings or something else?)
*ABE015:	*también, porque el audio me ha parecido un poco difícil* (also, because I think that the audio was too difficult).
*ABE027:	*yo sí cambiaria lo de escuchar* (yes, I'd change the listening part).

*ABE001 *sí, yo también* (yes, me too).
*RES: *lo de escuchar, ¿porque no se oía bien o…?* (the listening, because you couldn't hear well or…?)
*ABE015: *algunas cosas hablaban muy rápido* (some of the things went too fast).

(39b) FFI+Collab

*MAR029: *que la del audio hablara un poco más lento para entenderle mejor* ([I'd like] that the girl of the recording spoke more slowly to understand her better).

In (39a), first a learner from Comp indicates that he felt "good, very good" about the tasks, but right afterwards, one of his peers points out to the fact that he did not enjoy them because of the listening component. After this comment, a few other participants support this claim, and suggest that the audio recording was difficult to understand because it went too fast. This same remark is shared by another learner from FFI+Collab, in example (39b).

In the focus groups where learners from Comp and Collab were mixed, the discussion regarding the second question revolved mainly around the comparison between the individual and the collaborative setting. Some participants from Comp felt that carrying out the task in pairs was better: *"a ver, yo no estuve en parejas, pero creo que si estuviera en parejas sería mejor porque así tienes la ayuda del otro, no estás solo y no te sientes marginado"* (well, I wasn't in pairs, but I think that in pairs it'd be better because you can rely on your peer for help, you're not alone and you don't feel left out) [ABE057], *"yo trabajé individual, y bueno pues estuvo bien, pero yo creo que es mejor trabajar en parejas"* (I worked individually, and well, it was fine, but I think it'd be better in pairs) [ABE009]. Conversely, others considered that the individual mode was more efficient: *"pues me parece bien individualmente porque así estás… escribes tú mismo"* (I think individually is fine because this way you are… you write by yourself) [ABE001], *"porque igual el compañero te puede distraer"* [ABE027].

The collaborative mode was very well perceived by learners in Collab. However, as it was already shown in some of the responses in DQ from learners in FFI+Collab, these children referred occasionally to conflicts that arose during the DG task performance, as in the excerpt below:

(40)

*MAR024:	*bien, solo que a veces cuando yo digo algo y me dice que no, que lo hagamos como él quiere.* (good, except for the fact that, when I say something and he says no, we must do the way he wants)
*RES:	*¿y es difícil ponerse de acuerdo?* (and it's difficult to agree?)
*MAR024:	*hombre, a veces no pero otras sí, por ejemplo…* (well, sometimes no, but others yes, for example…)
*RES:	*¿mandaba mucho?* (was he very bossy?)
*MAR024:	*bueno, yo también, es que imagínate, yo le digo que escriba una palabra en inglés y me dice "no, se hace así", y puede que él a veces tenga razón pero otras la tengo yo, y cuando te preguntamos a ti y él tiene razón se pone como chulito.* (well, me too, just imagine, I tell him to write a word in English, and he goes "no, it's not like that", and he might be right, but others I'm the one who is right, and then we check with you, and he is right, and he starts to rub it in)

In (40), MAR024 describes a pair work situation where their disagreements were not managed adequately, as she mentions dominant attitudes by both sides. To conclude the analysis of their responses to the second question, as can be seen in (41), a learner in Comp indicated feeling stressed when carrying out the text reconstruction stage, which supports the fact that this stage was the least preferred part in the open-ended responses in DQ:

(41)

*ABE057:	*por ejemplo, estás escribiendo y no te sale una palabra y estás ostras, a ver cómo se dice, a ver cómo se dice… […] te ponen nervioso y aunque al principio no lo estás, luego te pones nervioso y al final lo pasas muy mal* (for example, you are writing and you can't come up with a word, and you go, shoot, how do you say it, how do you say it... […] they make you feel nervous and, although at the beginning you aren't, then you start feeling nervous and in the end you have a very bad time)

In excerpt (41), ABE057 claims to feel nervous whenever she cannot think of a word in English while writing, and refers to the fact that the feeling of anxiety increases progressively while performing the task.

The last question in the focus group interviews aimed to gauge child learners' opinion about the whole experimental procedure. In general, learners in all groups expressed positive feelings about our intervention: *"también me ha gustado porque así* *ya volaras también otras cosas, no sé, ves el inglés diferente"* (I also enjoyed it because this way you can appreciate more other things, I don't know, you perceive English in a different way) [ABE043]. However, a few learners also believed there were too many sessions: *"cansa"* (it's tiring) [ABE039], *"lo que no me ha gustado es que hiciéramos tantas, si hiciéramos una de cada, pero ya tantas veces…"* (what I didn't like was that we did too many… If we did one task of each… but so many times…) [ABE057]; *"más seguidos"* (more concentrated) [MAR014].

Last but not least, some learners also provided some suggestions for improving the experimental design: *"yo en lo de escuchar y escribir, yo cambiaría en el primero unas parejas y en el segundo otras, para comparar si trabajan mejor o igual"* (I for one in the listening and writing task I'd change the pairs from the first day to the second day, to see if they work better or the same) [MAR001]; *"en lo de las caritas, al final del todo dejar una especie de textito para que la gente pueda poner como qué opina de alguna pregunta"* (in the smiley thing [the attitude questionnaire], at the very end, a short text block so that people can tell what they think of a question) [MAR012].

We will now turn to the analysis of the third and last research question related to individual differences (RQ3c). In this question, we aimed to examine the extent to which learners' L1 writing skills (measured by means of a story continuation task in the L1) and attitudes towards writing (measured by one the subdimensions in the AQ, explained above) related to their L2 writing quality. In fact, previous research has found a correlation between L1 and EFL writing skills (Pae, 2018; Ströbel et al., 2020), in some cases, even more so at lower level of proficiency (Schoonen et al., 2011), and also between writing perception and EFL writing skills (Pae, 2008).

In our study, EFL writing was assessed through an array of text-based measures (tapping into complexity and accuracy), and also through an analytic rubric. For the current analysis, the quality of L2 writing will be considered in terms of the lexical diversity index (Guiraud's Index, GI) at T1, that is, before the DG intervention. In fact, lexical diversity has been demonstrated to be a strong predictor of EFL writing quality and EFL proficiency in general (Crossley & McNamara, 2012; Treffers-Daller et al., 2016). Moreover, in our data, the comparison of individual writing across the three time points also allowed to see that GI remained the most stable measure across the three experimental conditions (i.e., less affected by collaborative DG and TR), and therefore, it may constitute a better indication of children's L2 writing expertise regardless of other confounding variables.

Besides, learners' attitudes towards writing was obtained from the AQ subdimension (see above), and finally, L1 writing proficiency was gauged through a story continuation task in Spanish, administered on the first week of the experimental procedure. The quality of the L1 narration was assessed by means of an analytic rubric, which had been previously validated (Fernández et al., 2019).

In order to investigate this question, a multiple regression analysis was performed, in which there was one criterion variable (CV), namely, L2 writing quality (L2W), and two predictor variables (PV), more specifically, learners' writing attitude (WritingAtt) and learners' L1 writing quality (L1W). As we were interested in ascertaining the unique contribution of each PV to the model and there was no previous strong empirical evidence that suggested that L1W was more relevant than WritingAtt or vice versa, we opted for a standard regression analysis. The complete statistical output can be found in D11, in APPENDIX D.

We followed the procedure outlined by Jeon (2015) and Larson-Hall (2016). First, we performed a data screening to check that the data matched the assumptions to conduct this kind of analysis. The sample size was n = 71, which is well beyond the suggested benchmark of 15 participants per PV for a reliable analysis (Stevens, 1996). Besides, we checked the z-scores of all variables (CV and PV), and they all met the assumption of normality (+3/-3). Thirdly, the Mahalanobis test, which measures how far each case is from the multivariate mean, failed to show any departures from the chi-square value of $\chi^2 = 16.27$ (with $p < .001$). Finally, multicollinearity was checked through bivariate correlation statistics (see Table 66), as well as through the variance inflation factor (VIF). The Pearson correlation statistics showed that all variables were significantly correlated with each other: the CV was moderately correlated with the two PV ($r = .30 - .40$), and the two PV had a low correlation between each other ($r < .30$). In addition, VIFs were below .40, indicating absence of collinearity.

Table 66 Correlation matrix

	L2W	WritingAtt	L1W
L2W	1	.450**	.345*
WritingAtt		1	.221*
L1W			1

Note: **Correlation is significant at $p < .001$; *Correlation is significant at $p < .05$

Afterwards, the first standard regression attempt was made. First, the normality of the residuals was checked by means of a scatter plot with the standardized scores. It could be observed that the pattern met the normality requirement except for case 12, which was beyond the -3 benchmark and, therefore, was considered an outlier. We decided to leave this case out and repeat the analysis. This time, a closer pattern of normality was obtained in the scatter plot, as all cases were within +3/-3 values. The model obtained in the second round of standard multiple regression analysis can be found in Table 67:

Table 67 Regression model summary

	Total R^2	Adjusted R^2	Intercept	L1W	WritingAtt
B	0.27	0.24	0.98	0.37	0.62
95% CI			[-0.29 – 2.25]	[0.06 – 0.67]	[0.28 – 0.97]
relative import. (sr^2)				0.06	0.14

The total R^2 for this model was 27%, meaning that the inclusion of the two PV explained 27% of the variance in the L2W scores, the CV. This value falls close to the median R^2 found in L2 research literature (R^2 = 0.32) (Plonsky & Ghanbar, 2018). This standard regression analysis resulted in the unstandardized regression coefficients that are also displayed in Table 67. Looking at the CIs, we can observe that none of the two PV crossed zero and, therefore, they significantly influenced the dependent variable (L2W). Finally, the sum of sr^2 values, which show the unique contribution of each PV to the model, explain that 20% of all variance in the model. This means that 7% of the variance comes from the shared variance (27 – 20 = 7%). In this model, WritingAtt has a greater unique contribution to the model (14%) than L1W (6%), and hence, it constitutes the most important variable for predicting L2W.

6.4 Summary

In the following table, the most important findings of the three sets of research questions are summarized:

Table 68 Summary of the main results

RQ1	RQ1a: FonF on Day 1
	• Individuals (Comp) – Low number of wLREs, mostly addressing lexis and spelling • Pairs (Collab & FFI+Collab) – Most LREs related to mechanics and F-others – In terms of turns, Lex LREs were most numerous and lengthiest – Low number of LREs and turns related to the target forms
	RQ1b: Characteristics of LREs on Day 1
	• Outcome – Most LREs were correctly resolved (with the proportion being greatest in Mech) – In Lex, child learners tended to address linguistic problems, leaving them unresolved • Depth of engagement – Simple > Elaborate in all topic categories except for Lex, where the opposite was true • PKL vs L2 use – L2 turns > PKL turns in F-targets and F-others – L2 turns < PKL turns in Lex, Mech and Disc – In F-targets and Disc, there were significantly longer turns in PKL than in L2
	RQ1c: Impact of task-related variables (TR & pretask FFI)
	• wLREs (Comp) – Day 1 > Day 2. The topic of children's reflections did not vary (Lex and Mech) • LREs – Time on task: FFI+Collab > Collab; FFI+Collab Day 1 > Day 2 – Focus

	▪ FFI+Collab produced significantly more Lex LREs than Collab, regardless of the day ▪ Day 1 vs Day 2: little variation in the average of LREs and turns – Outcome ▪ Similar rates on Day 1 and Day 2. On average, higher number of correctly resolved LREs than other outcome categories – Engagement ▪ Elaborate LREs: FFI+Collab > Collab, regardless of the day ▪ Elaborate LREs decreased from Day 1 to Day 2 for all participants, but simple LREs increased – PKL vs L2 use ▪ FFI+Collab produced significantly longer turns in PKL than Collab, regardless of the day
	RQ1d: Attention to the target forms (3S & POSS)
	• Low number of LREs devoted to the target forms • There were significantly more LREs about 3S and POSS when these forms were targeted by DG1 and DG2 (respectively), than when they were not • In DG1, FFI+Collab produced significantly more 3S LREs than Collab
RQ2	RQ2a: Patterns of interaction
	• The collaborative pattern predominated on both experimental days • Pairs in Collab were more often described as collaborative than in FFI+Collab (where more dominant/dominant and dominant/passive patterns were observed) • Moderate impact of TR on these children's patterns. Changes in dyadic behaviors from Day 1 to Day 2 did not always imply becoming more collaborative

RQ2b: Analysis of the written production on Day 1 and Day 2 in response to the DG task

- Clause type
  - FullCRatio higher than the rest of the clause categories on both days (especially in Collab). Significant decrease from Day 1 to Day 2
  - PreCRatio and ProtoCRatio significantly increased from Day 1 to Day 2
- Complexity
  - Grammatical
    - Low use of subordination
    - Large use of coordination. CoordC/T-unit significantly decreased from Day 1 to Day 2, irrespective of the experimental condition
  - Lexical
    - Similar GI on Day 1 and Day 2 in all conditions
- Accuracy
  - Grammar: Learners working in pairs (Collab and FFI+Collab) showed a trend for becoming more accurate on Day 2
  - Spelling: the highest error rate corresponded to FFI+Collab on both days
  - Lexis: FFI+Collab highest borrowing rate
  - Content: average of 5-8 ideas recalled (with the highest rate corresponding to Collab)
  - Target forms: FFI+Collab highest accuracy (with large variability)
- Rubric
  - Collab obtained the highest mean scores in all dimensions except for Acc, where Comp showed an advantage
  - Collab scored significantly higher than FFI+Collab in Mech, regardless of the day

RQ2c: Analysis of the individual written production at T1, T2 and T3

- Clause type
  - Children produced a high ProtoCRatio, regardless of the experimental condition and the testing time
  - PreCRatio T1 > T2; In FFI+Collab, FullCRatio T1 < T3
- Complexity
  - Grammatical
    - Subordination < Coordination
    - C/T-unit: downward trend in Collab from T1 to T2

	– Lexical ▪ Clear differences between the three experimental conditions from the start (Collab showing the highest GI, and FFI+Collab the lowest) • Accuracy – Grammar: Collab showed the best performance. FFI+Collab displayed a downward trend from T1 to T2 – Lexical: the highest Borrow100 corresponded to FFI+Collab (but T1 > T2). Upward trend in Collab and Comp – Target forms ▪ 3S: FFI+Collab highest accuracy at T2 and T3. Presence of outliers ▪ POSS: large variability, FFI+Collab upward trend from T1 to T3 • Rubric – Collab's performance superior to that of FFI+Collab and Comp at all testing times – In Mech and Lex, Collab scored significantly higher than FFI+Collab
RQ3	RQ3a: Attitudes towards writing and collaboration (AQ results)
	• Comp and Collab displayed a significantly more positive disposition towards writing than FFI+Collab • In contrast, all learners were similar in their attitudes towards collaborative work • L2 writing triggered more negative feelings than L1 writing, especially in the case of FFI+Collab
	RQ3b: DG task perceptions (DQ and focus group interview results)
	• Learners who experienced the task in pairs (Collab and FFI+Collab) displayed more positive attitudes towards the DG task – Individuals mentioned more often the listening stage as one of their preferred aspects, whereas in the case of FFI+Collab and Collab they mostly referred to pair work – Writing and the text reconstruction ranked among the least preferred aspects for all participants • Collab displayed a significantly better perception of their own setting than FFI+Collab

	• Learners from FFI+Collab and Comp expressed a positive opinion about the pretask FFI stage and the individual languaging stage, respectively • In the interviews, some children claimed that they struggled to understand the audio recording • Some children in Comp complained that they were not given a choice to work collaboratively, as they considered this setting to be more fun and less anxiety-inducing • Conflicts in the pair work during DG were most often mentioned by learners in FFI+Collab • In general, children appreciated the chance to participate in the experiment as they considered that it provided them with a different approach to English and they enjoyed breaking the school routines
	RQ3c: Predicting L2 writing achievement
	• According to our statistical model, attitudes towards writing carry more weight than L1 writing quality for predicting L2 writing achievement • The sum of the two predictor variables (i.e., attitudes towards writing and L1 writing quality) explained 27% of all the variance in L2 writing scores

CHAPTER 7 DISCUSSION

In the present dissertation, we aimed to examine how YLs (aged 11-12) responded to a CW task, namely, a DG task, and to investigate the impact of this task on the learners' L2 writing quality. Moreover, we also set to analyze the interplay of various task-related variables, including TR and FFI, as well as certain individual differences, such as attitudes and L1 writing proficiency. In this section, the main findings will be discussed in relation to the three sets of research questions and their initial hypotheses.

7.1 Dictogloss, Focus-on-Form (FonF) and the impact of task related-variables

In the first set of research questions, the nature of the written self-directed languaging (wLREs) and the collaborative dialogue (LREs) of child learners during a DG task was examined in relation to a task design variable (with or without pretask FFI) and an task implementation variable (TR).

RQ1a What is the focus of YLs' written languaging (wLREs) and oral language-related episodes (LREs)?

First, the written reflections from child participants in Comp and the oral interaction of all pairs (that is, belonging to Collab and FFI+Collab) from the first DG performance (Day 1) were analyzed. The aim was to have a general perspective of how child participants in the individual and the collaborative settings attended to form in the task, isolated from the impact of task repetition (TR) and focus on form instruction (FFI).

Regarding the self-directed notes written by the participants in Comp, the findings showed that these YLs struggled to express in writing what linguistic aspects they had problems with during the reconstruction. In fact, a large proportion of their comments did not constitute deliberations about language, but rather comments on the task difficulty or their individual ability to carry out the task. Yet, when wLREs did occur, they mainly revolved around lexis and mechanics. Moreover, there was no episode related to the target forms (3S and POSS). These findings contrast with the results from Ishikawa (2018) and Ishikawa and Révész (2020), who demonstrated that the use of written languaging after an individual DG task was beneficial for low proficiency adult learners written accuracy, who were unfamiliar with the target form (the English past counterfactual) prior to the experimental procedure.

However, there is a key difference between their design and the one employed in the present study which is worth noting. Learners in Ishikawa (2018) and Ishikawa and Révész (2020) received the original DG text and were explicitly told to compare it with their own production, encouraging them to deduce the grammar rule of the target form. Therefore, the procedure used by these authors departed slightly from the original DG task procedure (García Mayo, 2018a; Wajnryb, 1990), as they introduced a stage which in the literature has usually been described as text modeling (Hanaoka & Izumi, 2012). However, research on the use of models by young EFL children has shown that this technique does not necessarily guarantee a greater attention to grammar, but rather it tends to promote lexical and content deliberations instead (Coyle et al., 2018; Coyle & Roca de Larios, 2014; Luquin & García Mayo, 2021).

In the current study, following previous research examining the DG task (Calzada & García Mayo, 2020a, 2021a; Kim & McDonough, 2008; Leeser, 2004; Malmqvist, 2005), learners attended a practice session involving an individual DG and wLRE in one of their Basque lessons. This session was led by their tutors before the experimental intervention took place. Contrary to Calzada and García Mayo (2020b, 2020a), the reason to opt for a language other than English was to avoid any potential impact of this training on the English or Spanish writing tests and target form encounters. Although this training may well have served to familiarize these child learners with the task procedure, it was probably insufficient to get them accustomed to languaging in writing, and even more so in EFL. Luquin and García Mayo (2021), who examined how YLs compared their own collaborative production with a text model, also pointed to the need of increasing the practice sessions prior to the pedagogical intervention in order to encourage their attention to grammar.

Notwithstanding the increasing evidence of children's capacity to establish crosslinguistic inferences and rely on explicit as well as implicit cognitive mechanisms (Roehr-Brackin & Tellier, 2019; Tellier & Roehr-Brackin, 2017), YLs' EFL previous learning experience might also affect the way in which they engage with languaging tasks in the L2. In fact, it has been shown that grammar is still presented in an implicit manner in

Primary School textbooks in Spain (Gris Roca, 2017). To conclude these comments on the ability of learners in Comp to reflect on language, we should not overlook the very likely possibility that they engaged in silent languaging while reconstructing the test (Cumming, 1990; Ishikawa, 2018). Whether or not externalizing those thoughts (as learners in Collab and FFI+Collab did orally) constituted an advantage will be examined later in the discussion on the set of RQ2s, when the written production assessment is considered.

As regards child learners' oral languaging in the first DG enactment, in line with previous studies that analyzed YLs' LREs during a DG focusing on 3S (Calzada & García Mayo, 2021a) and a DG focusing on 3S + Articles (Calzada & García Mayo, 2020a), our present findings confirm that these children were able to focus most of their attention on mechanics and grammar (well above 60% of all LREs). This result highlights the potential of DG as a form-focused task, despite previous claims that it would mainly trigger lexical LREs among low-proficiency learners (Leeser, 2004).

The choice to detach mechanical LREs (i.e. spelling and punctuation) from grammar focused LREs, which had been categorized together as "form" in previous studies (Alegría de la Colina & García Mayo, 2007; García Mayo & Azkarai, 2016), also allows us to observe some interesting findings, echoing those reported in Calzada and García Mayo (2020b). On average, children in the current study generated most LREs about mechanics (significantly more than the rest of the LRE focus categories). Yet, if the mean length of turn is considered, Mech goes down to the third position, right after Lex (the lengthiest LREs on average) and F-others. In the literature, it is now common to take into account not only the quantity of focus-on-form, but also its quality (Fortune, 2005; Zhang & Plonsky, 2020), and it has been suggested that more turns per LRE are an indication of more engagement (Storch & Wigglesworth, 2010). We will go back to the notion of engagement later in the discussion of RQ1b, where we analyzed LREs regarding children's "Simple" or "Elaborate" engagement, as well as with regards to their use of PKL.

There are two more aspects that are worth noting as far as the focus of LREs is concerned. Firstly, despite the fact that discourse LREs (Fortune & Thorp, 2001) have not usually been explored in process-oriented CW research (Storch, 2019a), we decided to tally them for their potential relevance when accounting for any changes in child learners' writing quality. However, on average, the participants in our study did not focus their attention on text-level aspects. Therefore, it should be questioned whether Disc deserves a separate category or whether pronominal reference searches or discussions about conjunctions should rather be included within F-others (Storch, 1997).

And secondly, although we will refer to them more in detail when we examine RQ1d, learners' deliberations about the target forms (3S and POSS) should be considered at this point. If compared to F-others, on average, target forms were discussed significantly less, as was the case in Calzada and García Mayo (2021a), and the raw proportion was virtually identical to the one reported in that study (around 6% of all LREs). Furthermore, the F-target mean length of turn was also second to last (only slightly longer than D-LREs), implying that these participants did not deliberate about the target forms in depth. However, F-target LREs also display the highest variability in data, which could possibly be related to differences between the two experimental conditions, since one of them, FFI+Collab, received pretask instruction on the target forms (discussed below in RQ1d).

In conclusion, the analysis of languaging on Day 1 reveals that learners in the individual condition had difficulties in reflecting metalinguistic issues in writing, and that when they did so, their comments dealt mostly with mechanics and lexis. On the other hand, the focus of the LREs shows that the DG task geared most dyads' discussions to mechanics and non-target grammatical features, although the length of the LREs indicated that these children discussed in more depth lexis and other grammatical features than mechanics. Finally, it was also observed that both the quantity and quality of LREs devoted to target features and discourse was lower than in the rest of categories.

RQ1b What is the LRE outcome, depth of engagement and Previously Known Language (PKL) use?

Having now considered the focus of LREs, we now turn to interpret the results of RQ1b, that is, how child learners resolved those LREs (outcome), to what extent they were invested in their discussions (engagement) and what languages they used for those deliberations (PKL use). Starting with outcome, as in Calzada and García Mayo (2021a), child learners in this research resolved most of their discussions correctly by relying only on their own resources in all focus categories. The average of correct LREs was especially high in Mech, which concurs with Agustín Llach's (2011) findings about L1 Spanish child EFL learners' mechanical skills. More specifically, this author reported that, despite the difference in the graphophonological systems of Spanish and English (shallow vs deep writing system, Cook, 2010), EFL primary learners by year 6 (ages 11-12) have already a stronger command of spelling than in year 4 (ages 9-10).

In contrast, the average rate of incorrect LREs, while being much lower than that of correct LREs, was slightly higher in F-others and Lex than in the rest of the categories. Although the average of unresolved LREs was low (similar to Calzada & García Mayo, (2021a), in the case of lexical discussions child learners produced significantly more addressed LREs than incorrect LREs. This differs from Calzada and García Mayo's (2021a) study, where addressed Lex LREs had still been minimal (0.09 on average per dyad). Nonetheless, the difference in proficiency (A1 - A1+ vs A2) and setting (mainstream EFL vs CLIL) between the children in the two studies could account for this divergence, as those in the current study may have been less familiar with the same DG text vocabulary, and hence, spent longer time trying to find the correct words without reaching any resolution. The fact that Lex episodes were left unresolved to a greater extent than other focus categories also coincides with Villarreal and Munarriz (2021).

Spending more time in linguistic deliberations is directly linked to the next LRE dimension: engagement. In fact, Lex LREs showed again a different pattern, as child learners produced on average significantly more elaborate than simple episodes, while simple engagement prevailed (not significantly) in the rest of the focus categories. Interestingly, looking at the use of previously known languages (PKL), children generated on average more turns and words in the their L1 than in the L2 when discussing lexis (as was the case in Disc and Mech). By contrast, there were more turns predominantly in the L2 when children discussed grammar (F-targets and F-others). Overall, these learners made a large use of their L1 during the DG task, accounting for more than 50% of the turns in Lex, Mech and Disc episodes.

These findings corroborate the results of recent studies on PKL use during child interaction in CW tasks (Martínez Adrián & Arratibel Irazusta, 2020), where YLs were shown to make an extensive use of their L1. Nonetheless, there are reasons to believe that the choice of the language does not depend so much on age as it does on task modality. In fact, while previous research on child EFL learners' interaction during oral only tasks (García Mayo & Hidalgo, 2017; Pladevall-Ballester & Vraciu, 2020) had reported a minimal use of the L1 during peer interaction, other research reported that its use significantly increased when the tasks included a writing component, both in the case of child (Martínez Adrián & Arratibel Irazusta, 2020) and adult learners (Azkarai & García Mayo, 2015; Payant & Kim, 2019; but see Storch & Aldosari, 2010).

A possible explanation for the preference of the L1 over the L2 in oral + writing tasks could be that, akin to individual writing, learners perceive these tasks as more product-oriented than process-oriented. Hence, they may believe that their goal is to produce an acceptable text in English, and make use of all their multilingual resources available to achieve that objective (Payant, 2020). The L1 would, therefore, fulfill and instrumental and mediating use, in accordance with sociocultural views of language (Storch & Aldosari, 2010).

In our study, child learners fell back on their L1 especially for lexical, mechanical and discourse deliberations. The preference for PKL when discussing vocabulary is aligned with previous studies where lexical searches accounted for most of children's L1 use (García Mayo & Hidalgo, 2017; Martínez Adrián & Arratibel Irazusta, 2020; Pladevall-Ballester & Vraciu, 2020). In this regard, there are reasons to argue that the more elaborate engagement in the case of Lex (discussed above) could be related to a greater PKL use. In fact, the length of Lex turns in PKL was significantly greater than that of L2 predominant turns. On the other hand, the lower amount of PKL predominant turns in F-others and F-targets could account for their more superficial engagement in these topics. Yet, it was also observed that when children discussed F-targets in PKL, those turns were significantly longer than the L2 predominant ones. Thus, PKL could also be associated with more in-depth languaging about 3S and POSS.

Furthermore, the fact that YLs resorted less often to their PKL for discussing grammar concurs with Martínez Adrián and Arratibel Irazusta (2020). Although these authors claim that their participants were less able to discuss grammar due to the nature of their instruction setting (CLIL), which tends to favor a communicative language approach, we have reported that the same holds true for mainstream EFL learners. In fact, primary school learners, although presumed to possess language-analytic abilities and developing explicit mechanisms of learning (Roehr-Brackin & Tellier, 2019), may still lack the necessary resources to verbalize certain grammatical aspects not only in their L2 (Macaro, 2005), but also in their L1.

At this point, a brief comment on the methodology followed to assess PKL in collaborative tasks should be made. As Azkarai and García Mayo (2015) and Martínez Adrián and Arratibel Irazusta (2020), we chose to classify turns as PKL or L2 depending on the predominance of the language (i.e. the highest number of words in one of the two languages). However, other scholars have opted for a classification based on the number of L1 terms per AS-unit (García Mayo & Hidalgo, 2017; Storch & Aldosari, 2010), and in some cases the criterium to quantify L1 use has not been reported (García Mayo & Imaz Agirre, 2016). Taking an example from García Mayo and Hidalgo (2017), "the T-shirt is *azul* (blue)" (p. 6), classified as an L1 use instance in their dataset, would have been classified as an L2 predominant turn according to our criteria. Therefore, these divergent methodological decisions could compromise the conclusions drawn across several pieces of research examining the same phenomenon.

RQ1c What is the impact of procedural task repetition (TR) and pretask focus-on-form instruction (FFI) on YLs' focus on form?

In the present study, child participants carried out a DG task twice with one-week interval in between. Moreover, some learners who were assigned to the collaborative setting (FFI+Collab) received a pretask focus on form instruction on the target forms (3S and POSS), while others followed the typical DG procedure (Collab). We will now discuss children's wLRE and LRE production in relation to these two independent variables: task repetition and experimental group condition.

With regards to wLREs, the average results indicate that child learners in Comp continued to show the same kind of reflections in the second DG enactment (Day 2).

That is, these participants expressed primarily mechanical and lexical comments on their individual writing, while wLREs on grammar and discourse were practically absent. Moreover, no reflection was made on either of the two target forms, but instead a great proportion of comments were devoted to their task perception and difficulties to recall the original text content. Thus, we can claim that repeating the task twice did not help these learners to become familiarized with written languaging, and it stresses the fact that more training is needed to focus their attention on linguistic issues.

Turning now to the comparison of the two experimental conditions that performed the DG tasks collaboratively (Collab vs FFI+Collab), we will first discuss the time they spent on the task. On both days, learners in FFI+Collab spent significantly more time than their counterparts in Collab on the DG tasks. It should be noted that participants in both groups were given the same amount of time to carry out the task (25 minutes maximum), and as can be observed, learners in FFI+Collab on average were more prone to using all their time than those in Collab. In this case, the differences in the setting may have played an influencing role. In fact, while dyads in Collab were withdrawn one by one to a separate classroom where the researcher or a research assistant supervised their task performance, dyads in FFI+Collab were taken five by five at a time to another classroom. Once there, the researcher provided the FFI to the whole group of learners, and right afterwards, during the collaborative DG stages, he circulated around the room. This second type of setting resembled more a real classroom atmosphere (i.e. was more ecologically valid) than the first setting, closer to a laboratory setting (Collins & White, 2019; Gass et al., 2005). In consequence, the FFI+Collab setting may have triggered a more spontaneous behavior from those child participants, as they may have felt less aware of being recorded or being involved in an experiment.

It should also be noted that the time that FFI+Collab spent on the task significantly decreased from Day 1 to Day 2. This finding is in line with previous research on the impact of TR on individual oral tasks with adults (Qiu & Lo, 2017) and collaborative oral tasks with children (Lázaro Ibarrola & Hidalgo, 2017). This decline could be interpreted as an indication that children were less engaged in the DG task (in terms, for instance, of what some scholars refer to as behavioral engagement, Philp & Duchesne, 2016), but, alternatively, it may also suggest that learners became more efficient at it (Qiu & Lo, 2017).

As far as LRE focus is concerned, as anticipated, TR did not influence the average production of LREs, which is line with previous findings (Hidalgo & García Mayo, 2019). Furthermore, the experimental condition did not trigger any significant differences in terms of LRE focus but, unexpectedly, learners in FFI+Collab produced significantly more lexical episodes than those in Collab. We hypothesized that the grammar activities on the target forms prior to the DG task would draw FFI+Collab's attention grammatical aspects , but as the average of F-others LREs showed, it was not the case. This finding concurs with Swain and Lapkin (2001) and Leeser (2004), who showed a little impact of another kind of pretask activity conceived to promote attention to form (more specifically, pretask modelling) on the nature of LREs during collaborative DG. The average of LREs related to the target forms produced by each experimental condition will be discussed below, in RQ1d.

A possible explanation for the higher average of Lex episodes in FFI+Collab might be again related to learners' vocabulary knowledge, already suggested above when discussing why in the present study there were more addressed L-LREs than in a previous investigation which examined the same task on a similar population (Calzada & García Mayo, 2021a). Although the difference in the *Flyers* proficiency test scores between the two collaborative conditions was not statistically significant, the difference in average values showed a medium effect size in favor of Collab. Consequently, there is a possibility that higher average results in the general proficiency test could imply a greater vocabulary knowledge, as has been pointed to in the case of low proficiency EFL learners (Terrazas Gallego & Agustín Llach, 2009).

When it comes to LRE outcome, there were more correctly resolved LREs than incorrectly resolved or unresolved in both experimental conditions, and resolution was unaffected by TR. In terms of engagement, FFI+Collab produced on average more elaborate LREs than Collab. However, this advantage could be influenced by the greater production of lexical episodes in the former condition than in the latter, as Lex tends to be more elaborate than grammar-related LREs. In this vein, it should also be noted that elaborate LREs decreased in FFI+Collab from Day 1 to Day 2 (although not significantly), whereas simple LREs increased in both groups. Therefore, the fact that FFI+Collab spent significantly less time on the task on Day 2 as compared to Day 1 (as mentioned above) could, indeed, be more related to a lower engagement than to a greater efficiency in the task management. Finally, regarding PKL use, FFI+Collab produced significantly longer PKL dominant turns than Collab, which is in line with two of the aforementioned aspects: (i) FFI+Collab's significantly larger production of Lex (associated with more PKL use), and (ii) their significantly larger production of elaborate LREs.

RQ1d Do YLs focus on the two target forms (3S and POSS)?

To conclude the discussion of the first set of research questions, we will refer to the extent to which (i) the especially designed DGs helped YLs focus on the target forms (3S and POSS) and (ii) whether providing pretask FFI made any difference.

When the DG text targeted 3S, on average, children generated significantly more LREs related to 3S than in the dictogloss which targeted POSS (and the opposite was true when the target of the DG was POSS). With regards to the impact of the FFI, those learners in FFI+Collab produced on average significantly more target LREs only in the case of 3S. Taking our previous DG studies into perspective, if these results are compared to the F-target LREs produced by the child learners in Calzada and García Mayo (2021a) (where they experienced the mainstream collaborative DG procedure), the mean difference with Collab is trivial (d = 0.05), whereas the difference in favor of FFI+Collab is large (d = 1.06). In other words, the inclusion of pretask FFI on 3S seems to positively affect child learners' attention to this form during DG. With regards to within group comparisons, no differences were found between learners' attention to 3S and POSS in either of the experimental groups.

We had anticipated that, due to the higher communicative value and salience of POSS (Sato & Loewen, 2018; J. White, 2008), this form would be more amenable to a DG task procedure and a FFI pretask stage with low proficiency L1 Spanish YLs, as compared to 3S. Yet, we did not find evidence for this advantage in our data. There may be three main reasons that account for the lack of significant differences in the attention to the two target forms. First, there were more target instances of 3S (n = 15) than POSS (n = 8) in the original DG texts. In fact, with the aim of keeping DG1 (3S) and DG2 (POSS) text length and genre as similar as possible, we had opted for lowering the number of POSS instances. Nonetheless, this decrease could have negatively impacted on the possibilities for languaging about POSS, even after a dedicated FFI on this form. The interplay between the number of target form instances in a DG and the possibilities for noticing them was already suggested as a potential moderator by Ishikawa and Révész (2020), whose individual DG task only contained three instances of the target form (the English past counterfactual).

Secondly, the contexts of POSS in DG2 were all circumscribed to family relationships. As previous research with Spanish-Basque learners of English had found (Imaz Agirre & García Mayo, 2013), more errors are usually made when the possessed entity is a human animate noun than when it is inanimate, and even more so when the learners' proficiency is elementary (Imaz Agirre & García Mayo, 2018). Moreover, the design of the family picture (see A3 in APPENDIX A), where kinship labels were added to all characters, as well as the proper names to the main characters, might have also negatively influenced learners' decision to use the possessive determiners. In fact, there were a few instances throughout learners' discussions and reconstructed texts (discussed below) in which there were omissions of POSS as well as definite article use (due to L1 transfer), as can be observed in the following example:

(42) [ABE007 & ABE011 - Day 2]

*CHI A:	her grandma…
*CHI B:	cards cards cards… cards! play with cards!
*CHI A:	*¡que no, tío!* (no way, dude!)
*CHI B:	grandmother…
*CHI A:	*¿entonces qué pongo?* (what shall I put then?) his grandmother?
*CHI B:	no, grandmother…
*CHI A:	pero hay que poner ↑su abuela (but it should be ↑his grandmother).
*CHI B:	¡no!
*CHI A:	*abuela juega a cartas, ¿así?* (grandmother plays with cards, like that?)
*CHI B:	the grandmother, *la abuela* (the grandmother).
*CHI A:	*pues hala* (ok then).
*CHI B:	the grandmother play with the cards.

In (42), when CHI A first mentions the character of the grandmother, CHI B is still focused on a different topic (play cards vs play with cards). However, right after insisting on his own choice (play with cards), CHI B goes on to question the use of the possessive determiner with ‘grandmother’, suggesting that this noun should go without it, but CHI A argues that they should clarify who the possessor is. Furthermore, CHI A translates the bare noun phrase to Spanish to show that it sounds inaccurate. Therefore, CHI B decides to add a definite article (the) before the noun, and it is finally then when CHI A compromises. This example serves to illustrate that, in general, these child learners’ command of POSS in the context of kinship relationships was still unstable. Although we did not aim to make a developmental assessment of the degree of acquisition of the target forms, in the case of POSS, following J. White’s (2008) classification, we could describe their use as being halfway between pre-emergence and emergence. Pre-emergence is characterized by avoidance of *his* and *her* or use of the definite article, while emergence is associated with a preference for one of the two forms. J. White (2008) also claimed that matched and unmatched kinship contexts (i.e. where the gender of the family relative matches the gender of the possessor or does not so) correspond to the later stages of the acquisition of POSS by English L2 learners.

Finally, another factor that could have affected the possibilities for noticing and discussing POSS in DG2 is related to learners’ previous instruction on these target forms. As the focus of the current study was attention to form, writing development and CW perceptions, participants’ explicit or implicit knowledge of the target forms was not gauged. Nonetheless, as indicated by Calzada and García Mayo (2020a), YLs in the educational context of the Basque Country are frequently provided at this stage with explicit instruction and corrective feedback on 3S. From the grammatical contents included in the English manual all participants in the present study shared (Read & Ormerod, 2018), there was a dedicated unit to family members in which possessive determiners were explained, whereas 3S appeared in different units throughout the textbook. In this vein, previous research in the Canadian ESL young classroom by Collins et al. (2009) found that POSS appeared rarely in teachers’ speech and classroom input (e.g. textbooks), and that it was even less frequent in the context of matched and unmatched family relationships. Therefore, apart from the differences in the amount of prior instruction, the sheer frequency of the target forms in the learners' mainstream EFL classroom environment may have played a role in their degree of focus on form during DG2.

7.2 Collaborative work

In the second set of research questions, we examined child learners' collaborative work during DG in a qualitative manner across the two enactments, by analyzing their patterns of interaction on Day 1 and Day 2 (RQ2a). We examined whether collaboration (with and without FFI) made any difference in terms of the writing quality of their reconstructed texts, as compared to individual writing (RQ2b). Finally, we also aimed to capture changes in YLs' individual L2 writing before and after the pedagogical intervention (RQ2c).

RQ2a What are YLs' patterns of interaction during collaborative dictogloss? Are they influenced by TR?

Previous research on YLs' patterns of interaction during DG had shown that dyads tend to display cooperative (or passive/parallel) and collaborative patterns almost to the same extent (Azkarai & Kopinska, 2020), while research employing other CW tasks and oral only tasks had reported mostly collaborative patterns (Butler & Zeng, 2015; García Mayo & Imaz Agirre, 2019). Our results align with the latter research body, as dyads in both experimental conditions (Collab and FFI+Collab), regardless of the DG day, were mostly collaborative (> 50%), whereas cooperative dyads constituted one third or less. Azkarai and Kopinska (2020) argued that the difference between their results and those by Butler and Zeng (2015) and García Mayo and Imaz Agirre (2019) could be due to the nature of the DG task, where one of the learners tended to play the role of "scribe" while the "non-writer took on some occasions the dominant role" (p. 10). In our study, having examined the same task type and a similar population (EFL learners in the BAC), we could not find evidence to support their claim.

One of the possible reasons could, in fact, be due to difference in proficiency within dyads. In Azkarai and Kopinska (2020), the *Flyers* test indicated that participants had an A2 level in English (somewhat higher than learners in the present study, who were described as A1 – A1+). Yet, no indication of the distribution of the proficiency scores was reported, and the pair formation method was not explained. Conversely, in the present study an attempt was made in order to keep the proficiency of the learners belonging to the same dyad as similar as possible. In fact, recent research on young EFL learners has demonstrated that even within lower proficiency groups, there may exist wide proficiency differences when participants are recruited from intact classrooms (Calzada & García Mayo, 2021c; García Hernández et al., 2017; Pladevall-Ballester & Vraciu, 2020; Vraciu & Pladevall-Ballester, 2020). In the case of CW, excessive proficiency differences within dyads have been claimed to hinder collaborative patterns (Choi & Iwashita, 2016; Kim & McDonough, 2008; Kowal & Swain, 1994), although other authors have found that its influence is moderated by other variables, such as collaborative mindset (Sato & Viveros, 2016). In fact, there is another possibility that child learners in Azkarai and Kopinska (2020) and those in our study differed in their predisposition towards collaborative work, a dimension that will be discussed below (in RQ3).

Regarding the impact of TR, informed by a previous piece of research on young EFL learners performing an oral task twice (García Mayo & Imaz Agirre, 2016), we had foreseen that the second DG would trigger more collaborative patterns. However, in general, the vast majority of dyads (21/31) did not experience any change in their dyadic pattern, whilst almost half of the dyads in García Mayo and Imaz Agirre (2016) experienced a change in the procedural TR condition. Furthermore, contrary to their findings, not only did our results reflect a positive development of dyadic interactions (i.e. showing higher mutuality and equality on Day 1 than Day 2), but also different kinds of shifts. In fact, the proportion of dyads that became collaborative on Day 2 (2/31) was lower than those who varied their pattern across lower mutuality and equality conditions (8/31).

It should be noted that, apart from using another task (i.e. an oral spot-the-differences task), there were two key differences between their study and the present one. On the hand, García Mayo and Imaz Agirre (2016) left a two-month gap between the two task enactments, while there was only a one-week gap in our study. On the other hand, their participants were somewhat younger (8-10 years old). Thus, it could be the case that younger children in a longer time span were subject to cognitive and behavioral changes to a greater degree than older children in the lapse of two consecutive weeks (Pinter, 2011). Furthermore, our findings are similar to those of Butler and Zeng (2015) and García Mayo and Imaz Agirre (2019) in the fact that, by ages 11-12, children display mostly collaborative patterns.

Why some dyads shifted their pattern on the second day could be due to several reasons, which in the absence of an immediate posttask interview or questionnaire (Wenxue Chen, 2018) remain totally speculative. For instance, in the case of ABE008 and ABE013 (a dyad from Collab), the fact that they were already familiar with the DG task procedure on Day 2 triggered a very strategic behavior from the two learners. During the listening stage, they divided the original text in two halves and each was responsible for the note-taking of their own part. Thus, in the reconstruction stage, they swapped the role of scribe and information provider and, without engaging in much languaging, they finished the task much faster than on Day 1 (17 min vs 11 min). Nonetheless, in other cases, such as ABE052 and ABE064, it could be observed that, while on Day 1 they felt very insecure, the second enactment allowed them to gain confidence. Below, the first minute of oral transcription is shown from this dyad on Day 1 and Day 2:

(43) [ABE052 & ABE064]

(43a) Day 1

*CHI B	eh... you write?
*CHI A	[answers with gestures to say that she should wait]
	[42 seconds quiet writing notes in the notes paper]
*CHI B	*¿ya?* (now?)
*CHI A	[whispering]
*CHI B	[whispering]
*RES	speak up, don't whisper ok? I'm not a teacher, don't worry, I'm just here to help Asier so you can speak freely.
*CHI B	I can speak… speak in Basque?
*CHI A	yes, yes, don't worry.
*CHI B	bale, *orduan kakahuete nola esaten da kakahuete?* (ok, then how do you say peanut?)

(43b) Day 2

*CHI B: you're ready? [CHI A only gestures] ok, celebration.
*CHI A: Tom, Tom…
*CHI B: yes but… eh… the celebration is in the garden, no?
*CHI A: yes.
*CHI B: eh… [starts writing] (…) ok so…
*RES: speak up ok? *altu hitz egin* (speak up).
*CHI B: ok.
*CHI A: [whispering] *pon punto* (put full stop) [normal voice] Tom is, is… playing football…
*CHI B: is playing football with…
*CHI A: a…
*CHI B: with an uncle.
*CHI A: yes [whispering] with her uncle.

In (43a), the learners are extremely inhibited to speak, and the research assistant has to encourage them to speak freely in whatever language they prefer. The dyad spends almost the whole first minute completely silent and there is no instance of collaborative dialogue except for a lexical search which was addressed to the research assistant. Conversely, in (43b), the learners appear more relaxed, especially in the case of CHI B, who seems to take the initiative. Moreover, although the research assistant still has to call their attention to ask them to raise their voice, the learners engage in languaging right from the first minute (*"pon punto",* "with her uncle"). Their time spent on the task also reflects this positive change: from only 6 minutes on Day 1 to 10 minutes on Day 2.

Finally, as far as the influence of FFI is concerned, we had not anticipated any impact of this pretask stage on the learners' patterns of interactions. However, our findings allowed us to observe that, despite the fact that the collaborative pattern predominated in both experimental conditions, the proportion in FFI+Collab was somewhat smaller (71% in Collab vs 53% in FFI+Collab). Once again, a potential moderating impact of the setting could be considered. The fact that the context in FFI+Collab resembled more a real classroom environment might have made these learners feel less constrained by the presence of the researcher and video cameras, and hence, triggered a more spontaneous behavior, occasionally leading to uncollaborative patterns. Although the lack of control over task management has been described as one of the drawbacks of implementing TBLT in young EFL classrooms (Carless, 2003, 2004), we provide ample evidence that in more than half of the cases all learners in the dyads were engaged in the task, and that the researcher was able to run the experimental procedure smoothly.

RQ2b To what extent do collaboration, TR and FFI influence the quality of the dictogloss written product?

In order to assess child learners' DG text reconstructions, a series of text-based and rubric measures were employed. Based on previous literature (Basterrechea & García Mayo, 2013; Fernández Dobao, 2012; Gallego, 2019; Shehadeh, 2011; Storch, 2005; Villarreal & Gil-Sarratea, 2019), we expected that children working collaboratively (Collab and FFI+Collab) would produce better quality writing than those working individually (Comp). Nevertheless, we did not find a clear advantage for the collaborative mode either in complexity (lexical or grammatical) or accuracy (grammatical, mechanical or content-related).

A very recent meta-analysis of CW studies (Elabdali, 2021) indicated that content-reproducing tasks such as DG showed fewer gains of the collaborative mode over the individual as compared to content-generating tasks (i.e. where learners are required to develop content of their own based on a prompt, such as graphs or essay topics for discussion). Our findings support Elabdali's suggestion, and in fact, it seems reasonable that if learners are given a model text to be reconstructed, their possibilities for using different linguistic structures or displaying their own writing styles will be more restricted than when they deal with more flexible tasks. A different question is whether in a more open narration task, such as the ones participants had to carry out as pre-, post- and delayed posttests, this lack of differences is still sustained (discussed below, in RQ2c).

Regarding the impact of TR, in general there were few significant changes between Day 1 and Day 2. As to the clause type child learners generated in their reconstructions, although on average full-fledged clauses (i.e. clauses that were contextualized with the task and accurate) dominated by far on both days, there was a significant decrease from Day 1 to Day 2, regardless of the experimental condition. In contrast, the protoclause ratio (i.e. clauses that were ungrammatical but contextualized with the task) significantly increased. The rubric dimension of accuracy also reflected a very slight non-significant decline in the average of accuracy scores, while adequacy remained unaltered. Nevertheless, the two text-based measures of grammatical accuracy (GramErr100 and GramErrFreeC) showed a different picture. In fact, there was a significant decrease of grammar errors per 100 words, which corresponded exclusively to the collaborative DG groups (Collab and FFI+Collab), and the average number of grammar error free clauses also increased (but not significantly), especially in the case of Collab. It follows from these results that impressionistic measures of accuracy, such as Torras' (2005) clause type classification or the rubric scores, do not always match those solely based on the quantification of text features. In fact, this disagreement was already signaled by Hidalgo and Lázaro Ibarrola (2020), who made a claim for the use of measures other than CAF to better gauge children's writing quality. However, it should also be noted, as pointed out by Elabdali (2021), that CAF measures are not capturing changes in low proficiency learners' writing may be derived from not employing suitable dimensions for this population.

This comment is equally relevant for discussing the next text-based dimension: grammatical complexity. Mirroring the literature assessing adult L2 writing, recent studies examining early L2 writing have used subordination indices (such as clauses/t-unit or number of dependent clauses) as the only indices of complexity (Bueno-Alastuey & Martínez de Lizarrondo Larumbe, 2017; Hidalgo & Lázaro Ibarrola, 2020). However, in the present dissertation a subordination index (C/T-unit) was employed, as well as a coordination index (Coord/T-unit), as has been suggested for beginner levels in English (Bulté & Housen, 2012; Mylläri, 2020). In fact, child participants were not expected to use much subordination, as the original DG texts they had to reproduce to did not contain these linguistic structures. What is more, the subordination measure failed to show any significant change.

Conversely, there was a significant decrease in the average number of coordinated clauses from Day 1 to Day 2. Although this decline could be understood as an indication of fewer complex structures in the second DG enactment, given the large use of coordination by these child EFL learners (on average, representing at least half of the T-units in their reconstructions on Day 1), it may be understood instead as an example of improved writing. Furthermore, if coordinated clauses decreased but subordination did not rise, it may actually mean that learners were using more simple independent clauses, and thus, segmenting their ideas in their writing more appropriately by means of punctuation. The two examples below, both from the same dyad [ABE007 & ABE011], serve to illustrate this change (errors are kept as in the original texts):

(44)

(44a) Day 1

Nora every days prepare her notebooks to go to scool // but her father forget prepare his sandwich // and Nora take money to buy apples on the supermarket // but on the supermarket Nora isn't buy apples. She buy black chocolate, // and she goes to scool with the face red // and when her teacher look her face // and say to go to the hospital.

62 words, 8 independent coordinated clauses

(44b) Day 2

Today is an a very good day, / Tom play football with his uncle. /The grandmother play with the cards. / Then is time to have lunch. / All the family was in the table to have lunch in the garden. / María ask to mothe [if she can take photos with her camara]. The afternoon is very hot // and they prepare some ice-creams. / Is a very beautiful day.

65 words, 6 independent simple clauses, 2 independent coordinated clauses, 1 dependent clause

In (44a), this dyad makes a great use of coordination to join the ideas, stringing together up to four such clauses in a row. In fact, there is only one full stop in the whole text. Conversely, in their second DG reconstruction (44b), while the text length is kept very similar (68 vs 71 words), these child learners decide to use simple clauses to a greater extent than coordination. This preference also comes with a more adequate use of full stops (six in total). If we relate this finding to our previous discussion about the oral interaction analysis (in RQ1), we could tentatively suggest a positive impact of learners' numerous LREs on mechanics (which were significantly more frequent than the rest of the focus categories) on the development of their use of punctuation in the second DG text.

In addition to linguistic quality indicators, in the present dissertation we also included a DG task content accuracy dimension, operationalized in the form of Idea Units (IU) (Carrell, 1985) retrieved from the original texts. Previous research had found that learners recalled significantly fewer ideas when working on a DG individually than in collaboration (Shin et al., 2016). However, we did not find support for this claim in our results, as there were not significant differences between learners in Comp and those in Collab and FFI+Collab. On average, learners in Collab retrieved more ideas than their counterparts, reaching at least seven IUs on both DG days. In other words, learners in Collab were able to recall at least half of the original IUs, whereas those in Comp and FFI+Collab were below that benchmark (ranging a mean of 5-6 IUs on both days).

The fact that these child learners were able to retrieve on average half or fewer of the original IUs could indicate, on the one hand, that they struggled to understand the content (due to unfamiliar vocabulary, for instance), or on the other hand, that they concentrated most of their effort on form and accuracy. Although in the current study we did not quantify content-related discussions (as opposed to what a handful of previous studies had done, such as Yang & Zhang, 2010), learners' collaborative dialogue was seeded with instances in which they also strived to obtain a text which was close to the original one. Example (45) serves to illustrate such discussions:

(45) Example of a content-related discussion [MAR008 & MAR017 – Day 2]

*CHI B	the mum is…
*CHI A:	the mother… *y si decimos* (and what if we say) her mother took a picture *¿o no lo decimos lo de que ha sacado fotos?* (or we just don't say that she took the pictures?)
*CHI B:	*es que, a ver, lo dice pero no sé qué dice* (I mean, the thing is that it says so, but I don't know what it says).
*CHI A:	*tenemos que decir más o menos que ha dicho, no todo* (we have to say more or less what the text said, not everything).
*CHI B	*ya* (I know).
*CHI A	*ahora hacemos hasta el final* (now we have more or less until the end) her grandfather… their grandfather… [starts writing again] (…) *eh, ¿qué mas han dicho?* (what else did they say)

*CHI B: *es que no se, ya al final que ha sido una gran celebración o algo así* (I don't know, at the end that there was a great celebration or something like that).
*CHI A: aha *sí* (yes) [writes the sentence] *hala, es que tiene que ver algo con estas fotos pero yo no he escuchado mucho* (that's it, it had something to do with these photos, but I didn't catch much of it).
*CHI B: *ya…* (I know…)
*CHI A: *si no hay aquí, vamos a inventar cosas* (if we can't come up with it we can invent things).
*CHI B: *¿se lo damos ya?* (shall we hand it in?)
*CHI A: *vale* (ok).

In (45) above, learners are stuck at the point in which they have to refer to what María's mother did in the story. They both agree that it was related to taking pictures, but they don't know exactly what the character did with the camera (in fact, it was a point in the story where two other characters, María and his father, took part). CHI A shows a more pragmatic approach to the task, reminding his peer that they are not required to write down all the narration, but only the gist of it. However, CHI B makes a further effort to try to recall the original message, by pointing to the drawing in the note-taking photocopy (conceived to help learners remember the story). In the end, when CHI suggests inventing ideas, CHI B prefers to turn the text in with what they already have, rather than making an attempt which would distort the original text. Finally, CHI B agrees to that.

This example shows that child learners in the present study were highly committed to conveying a coherent text which resembled the story they had listened to, and that, therefore, it is not likely that they had a completely grammar-focused approach to the DG task. Hence, it seems more reasonable to believe that the low number of IUs contained in the children's reconstructions is more related to a difficulty in comprehending the aural text than to their linguistic or form-focused priorities during the task. Calzada and García Mayo (2021c) found, indeed, that the amount of IUs retrieved in individual DG was moderated by learners' L2 proficiency. Furthermore, as will be discussed in RQ3b, learners also mentioned in their open-ended responses in the post-task questionnaire (the DQ) that they struggled to understand the audio recordings. Apart from L2 listening comprehension, there is also a possibility that the task imposed an excessive memory load to these child learners. In fact, it has been shown that YLs' memory capacity differs qualitatively from that of adult learners (Butler, 2016) and memory issues have also been demonstrated to influence YLs' perceptions of language test difficulty (Cho & So, 2014).

With regards to the reconstructed text quality differences between Collab and FFI+Collab, we had anticipated a positive effect of the pretask FFI on this group's written accuracy. However, we could not find any significant differences between the two groups in terms of general grammatical accuracy. Yet, the analysis of the rubric scores yielded a significant difference between the two groups in terms of mechanics. However, it should be noted that the rubric descriptor for "mechanics" only addressed spelling accuracy, and not punctuation. Therefore, this finding should not be interpreted as conflicting evidence with what has been discussed above in terms of the positive influence of mechanical LREs on children's writing.

Nonetheless, FFI+Collab did show a more accurate use of the target forms 3S and POSS than Comp and Collab, although the difference did not reach statistical significance. This result is in line with their attention to these forms, as FFI+Collab generated more discussions about the DG target features than their counterparts in Collab (a difference that reached significance in the case of 3S). Yet, this finding should be taken with a word of caution, as the mean accuracy score remained very low across all groups (below 50%). Moreover, although FFI+Collab showed more accurate production of the target form, they were also the group that produced on average the fewest obligatory contexts for both target features in their writing.

To conclude addressing RQ2b, a comment on the relationship between learners' LREs (discussed above) and their impact on the reconstructed text should be made. In fact, if languaging about language-related issues constitutes "language learning in progress" (Swain & Lapkin, 1998, p. 321), we would expect to find a connection between what learners speak and what they write. Studies in the CW literature have traditionally adopted either a process-oriented approach (i.e. examining LREs, patterns of interaction or engagement) or a product-oriented approach (i.e. focusing on the quality of que written production) (Elabdali, 2021; Zhang & Plonsky, 2020), but there are still few examples in the literature that have attempted to amalgamate both. As mentioned above, we could find some tentative evidence for the connection between LREs about other grammatical forms (F-others), and an increase in general grammar accuracy measures in Collab's and FFI+Collab's writing. Likewise, regarding target forms, not only did FFI+Collab generate more F-target LREs, but also yielded the highest 3S and POSS accuracy rate. This corroborates Basterrechea and García Mayo's (2013) finding in the CLIL adolescent setting, where more instances of LREs focusing on the 3S were positively correlated with higher written accuracy.

Moreover, the high number of mechanical LREs could also be related to more adequate punctuation and, therefore, more appropriate segmentation of the texts into sentences (reflected in the significant decrease of coordinated clauses). Yet, the numerous mechanical LREs did not translate into more accurate spelling, in line with Villarreal and Munarriz Ibarrola (2021). Neither did FFI+Collab's significantly higher average of lexical episodes translate into more lexical diversity in comparison to Collab. A lack of correspondence between the discussions generated during writing and the written production has also been attested in the collaborative DG literature. For instance, Kuiken and Vedder (2002a), who failed to find a connection between adolescent and adult FL learners' writing strategies and the grammatical complexity and lexical richness of the reconstructed texts, pointed to the possibility that the "discussion of a certain linguistic item (e.g. the correct use of a preposition) may lead to an increased awareness in other areas (for instance, word order)" (p. 186). Whether collaborative dialogue influenced YLs' individual writing or not will be discussed in the next question.

RQ2c Does dictogloss have an impact on subsequent individual writing?

In the last question of the second set of RQs, we adopted an experimental design approach to the analysis of the impact of collaboration, according to Elabdali's (2021) definition. In this type of studies, the emphasis is on the extent to which CW supports L2 learning and writing, and on determining whether that learning (if any) is sustained over time. Thus, we analyzed children's written output in the story continuation tasks they had to complete as pre-, post- and delayed posttests (referred to as T1, T2 and T3 in the present dissertation).

From the analysis of the results, several inferences can be drawn. Regarding the type of clauses, the DG task seems to have positively influenced learners' type of clauses. In fact, child participants, regardless of their experimental condition, produced on average significantly fewer preclauses (i.e. ungrammatical and uncontextualized clauses) from T1 to T2. Moreover, the ratio of full-fledged clauses (i.e. grammatical and contextualized clauses) progressively increased in the case of FFI+Collab throughout the three time points, and indeed, this ratio was significantly higher at T3 than at T1.

Observing the text-based measures, there was another indication that FFI+Collab's writing became more grammatically accurate after the DG intervention. Although a significant difference in the scores could not be found, this group's average ratio of grammar errors per 100 words decreased from T1 to T2, but increased again from T2 to T3. Moreover, the average of error free clauses increased from T1 to T2, and they were able to keep this gain at T3. As far as 3S target form production is concerned, FFI+Collab was the only experimental group that increased their accuracy score from T1 to T2, but it slightly decreased from T2 to T3. Last but not least, the accuracy dimension from the analytic rubric also only showed an improvement from T1 to T2 in this same group (and a slight decline from T2 to T3). In other words, there seems to be a trend from FFI+Collab to achieve better grammatical accuracy scores right after the DG intervention, as compared to Comp and Collab, who kept these measurements more stable across time.

In the rest of the text-based and rubric dimensions, we failed to see a clear pattern of L2 writing development as a result of the dictogloss task. In this vein, except for the case of general and target form accuracy in FFI+Collab, the fact of having experienced the task individually or collaboratively did not seem to make any difference in the rest of writing dimensions. Therefore, it could be the case that it was adding a pretask FFI to the DG tasks that influenced this group's production, and not the DG task per se. Yet, in order to find more support for this claim, future studies should include an individual DG condition with pretask FFI.

To conclude, we will now comment on different aspects related to learners' story continuation writing quality. Firstly, Collab's average rubric scores were significantly higher than FFI+Collab's in a number of dimensions, regardless of testing time, including adequacy, coherence, mechanics and lexis. Moreover, the significant difference in vocabulary was also observed in the text-based lexical diversity index (Guiraud's index). Furthermore, in the analysis of the DG reconstructions (discussed above in RQ2b), Collab also obtained significantly higher average scores in the rubric dimension of mechanics. Therefore, we could hypothesize that the fact that Collab outperformed FFI+Collab in these areas could imply that the former had an advantage over the latter from the start, in line with their higher proficiency scores obtained before the experimental procedure took place.

Secondly, if we draw a comparison between learners' DG text reconstructions and their story continuation production at the three time points, we can observe that, regarding clause type, learners in all experimental conditions generated on average more full-fledged clauses in the reconstructions than in the writing tests. In fact, in the writing tests, the average proportion of protoclauses was higher than that of full-fledged clauses at all testing times. Furthermore, the average of preclauses (i.e. ungrammatical and uncontextualized clauses), which was trivial in the reconstructions (representing less than 10% of all clauses on average), was higher in the writing tests (ranging from 5%-33%). Consequently, it could be argued that the type of task may have played a role, as the DG task falls under the category of content-reproducing tasks, while the story continuation tasks belong to the content-generating group. In other words, when a model was provided (i.e. the original DG texts), these YLs found it easier to produce clauses which were both grammatical and contextualized (full-fledged clauses) than when they lacked such resource in the story continuation tasks.

From this overall picture, it also follows that the category of paraclauses (i.e. grammatical but uncontextualized clauses), in line with Coyle and Roca de Larios (2014) and Torras (2005), was not relevant for the study of these children's written output, as the average proportion did not exceed 5% in any case. In fact, both the DG and the story continuation tests, although differing in the amount of content input provided, are both contextualized, and hence, it is more unlikely that learners will produce paraclauses.

Thirdly, the difference in task type can also explain why the number of POSS target form obligatory contexts was lower in the story continuation tasks than in the DG (max. 40 vs 145). In fact, the writing prompts provided at the tests failed to trigger the use of his/her, and therefore, how DG2, which targeted the use of POSS, affected the development of the written accuracy of these forms could not be determined. In contrast, the writing tests did succeed in generating obligatory contexts for the 3S to a similar proportion as DG1, which targeted 3S (max. 146 vs 145). This similarity could probably be due to the fact that the story continuation task already provided the beginning of the narration in the present tense.

Finally, if these findings are observed in parallel to learners' languaging in the DG tasks (discussed above, in RQ1), grammatical accuracy appears to be the only dimension in which a growth in the mean scores in the case of FFI+Collab's could be related to their grammar-focused LREs, which were on average most frequent after mechanics and lexis. Nonetheless, neither in Collab or FFI+Collab did mechanical or lexical LREs significantly affect their individual writing quality in these dimensions. Calzada and García Mayo (2021c), in a similar experimental design to test the impact of a collaborative DG on individual DG, could not find a correlation between learners' oral discussions and their writing gains. In their study, grammatical complexity (measured by means of MLC) was found to be the only dimension which significantly increased after experiencing the collaborative dictogloss procedure. The authors contended that more words per clause could be related to more use of subordination and coordination, and hypothesized that it could be the learners' LREs about grammar and mechanics that triggered that development. In summary, both Calzada and García Mayo (2021c) and the present study found some evidence of grammar development (either in terms of complexity of accuracy) as a result of having experienced a DG task.

7.3 Individual differences

Learning a foreign language is far from being a linear process (Huang et al., 2021), and hence, variability is intricately associated with SLA data (Tagarelli et al., 2016). Moreover, some researchers argue that child ISLA is generally more subject to individual variation, given that YLs' sociocognitive behavior and language are not as defined as that of adult L2 learners' (Oliver et al., 2017). In this respect, a number of individual differences (IDs) have been considered to influence L2 learning. In the present dissertation, variability was present in the findings from oral LREs (some pairs producing more metalinguistic discussions than others), pair dynamics (as some dyads failed to display collaborative patterns) or in the quality of their written production (for example, some learners obtaining higher accuracy and complexity rates than others). Thus, in an attempt to determine potential moderating factors of the effectiveness of the DG procedure, in the third set of research questions it was decided to study the following IDs: YLs' attitudes towards writing and collaborative work (RQ3a), YLs attitudes towards the DG task (RQ3b) and the impact of L1 writing skills and attitudes towards writing on L2 writing.

RQ3a What are YLs' attitudes to writing and collaborative work?

Research on CW tasks examining learners' affect towards this mode of working have generally concentrated, with a few exceptions (Kopinska & Azkarai, 2020), on tapping learners' beliefs after a certain experimental procedure (Calzada & García Mayo, 2020b; Fernández Dobao & Blum, 2013; Gallego, 2014; Lin & Maarof, 2013; Shak, 2006; Shehadeh, 2011; Storch, 2005). Yet, research from the psychological strand has highlighted the need to understand preconceived learners' disposition towards the L2 learning process (Sato, 2020b), as it has been shown to be a mediating factor in peer interaction behavior, influencing, for example, the amount of attention to form during oral tasks (Dörnyei & Kormos, 2000) or peer feedback provision (Sato & McDonough, 2020). This "initial condition" (Dörnyei & Kormos, 2000, p. 281) has been operationalized in various ways in the literature, including learner beliefs (Sato & Ballinger, 2016) or interaction mindset (Sato, 2017). In the present study, taking the two main features of CW tasks into account, we decided to tap into YLs' attitudes towards writing (a novelty in the early CW literature) and collaboration by means of a questionnaire (referred to as AQ) one week before learners' first encounter with the DG task.

Our findings showed that child learners displayed a positive attitude towards both writing and collaborative work (with an average that was above 3 in all experimental conditions for both dimensions) and, indeed, these results constitute further evidence about the suitability of CW tasks for this population. However, child learners in FFI+Collab showed significantly lower scores in their attitudes towards writing than their counterparts. In fact, it should be remembered that FFI+Collab's proficiency scores were also somewhat (but not significantly) lower than those in the other two conditions. Therefore, we could claim that FFI+Collab started the experimental procedure with a disadvantage as compared to Comp and Collab. Moreover, their less favorable attitudes towards writing could in part explain their worse scores already obtained at T1 (pretest, discussed above). At that stage, FFI+Collab performed worst in terms of type of clause (producing, on average, fewer full-fledged clauses and more preclauses), obtained the lowest text-based grammar accuracy ratios and widest use of borrowings, and they also obtained the lowest mean scores in all rubric dimensions. However, the possible correlations between writing scores and attitudes need still to be tested statistically (by using, for instance, correlation analysis or multiple regression).

Interestingly, however, all three experimental conditions displayed similar attitudes towards collaborative work, despite the fact that FFI+Collab (from School B) reported working more often in pairs than Comp and Collab (both from School A). Therefore, this lack of differences may indicate that, regardless of their prior experience, children possess a natural inclination towards collaborative work settings and tasks involving peer interaction (García Mayo, 2021c). Furthermore, these results go in line with Kopinska and Azkarai's (2020) previous study on YLs' attitudes towards this mode before the completion of a DG task.

Children's open-ended responses in the AQ provided deeper insights about their initial disposition towards L1 and L2 writing, as well as towards individual and collaborative work. Regarding writing in the two languages, irrespective of the experimental condition, these learners displayed rather opposing feelings about writing in Spanish and in English, with the former concentrating far more positive views than the latter. This finding corroborates what Calzada and García Mayo (2020b) reported in their exploratory study about young EFL learners' disposition towards these two dimensions. In contrast, the lack of correspondence between L1 and L2 writing perception is at odds with Saeli and Cheng's (2019) results in the Iranian adult EFL context. In fact, they argued that their participants' negative views of EFL writing were motivated by their previous experience in the L1. Our child participants' views are also somewhat different from those reported in the Hong Kong EFL primary context, where Shen et al. (2020) found a moderate motivation from learners aged 9-12 towards L2 writing, which, indeed, tended to be higher for those in the 6th year (aged 11-12). The authors argued that the increase in motivation could actually be due to the greater EFL writing practice these learners received as a preparation for the national exam to enter secondary education.

To understand these apparently mixed findings, the characteristics of the EFL context where the present study is set should be noted. In Spain, the teaching of EFL in primary school is usually characterized by a heavy reliance on oral communicative activities and tasks, while writing has been perceived as a mere support for vocabulary and grammar instruction, and relegated to homework activity (Mateo Cutillas, 2016). Conversely, by this year, learners' literacy skills in the L1 have been consolidated and their L1 writing can be considered mature. In fact, process-oriented writing activities are numerous in the most popular Spanish language textbooks of 6th year of primary school (Núñez Cortés & Barrado Mendo, 2020). Therefore, children with a low proficiency in English (i.e. in the range of an A1) and with little practice time spent on EFL writing may feel less confident about their L2 writing than about their command of L1 writing. Nevertheless, a different scenario emerges when these learners' views on collaborative and individual work are observed. Although the former attracts more positive opinions than the latter, the difference is not so great. In fact, children appreciated in general both modes of working (50% of their views were classified as positive in both cases, regardless of learners' experimental condition). This finding agrees with Kopinska and Azkarai's (2020), as they reported that YLs valued individual work because it allowed them to make their own decisions and stay more concentrated on the task.

In summary, the findings from the AQ indicated that child participants showed an overall positive initial disposition towards collaboration and writing, which provides support for the use of CW tasks in this setting. Nevertheless, learners in FFI+Collab showed less positive attitudes towards writing than the other two conditions, and when L1 and L2 writing beliefs were juxtaposed, all learners' tended to be less favorable about the latter.

RQ3b What are YLs' insights about the dictogloss task?

In this question, we sought to gauge learners' views on the DG task by means of a questionnaire (referred to as DQ) and focus group interviews conducted after the intervention. In general, the average attitudinal scores reflected these child learners' positive perception of the different aspects surrounding the DG task, since all the dimensions tapped by the DQ were above the 2.5 benchmark (in a 1-5 Likert scale). This favorable view about the task is in line with what Calzada and García Mayo (2020b) had already pointed in their exploratory study. Furthermore, as anticipated in our hypotheses, learners who had experienced the DG procedure collaboratively showed a more favorable opinion of the common DG task features, such as the listening, note-taking and the text reconstruction stage. This finding mirrors Kopinska and Azkarai's (2020), who, in a within-subjects design study, reported that YLs were more motivated in pairs than individually both before and after carrying out the task.

As far as learners' perception of TR is concerned, learners' displayed a general inclination towards having a second encounter with the DG task, in line with Shak's (2006) claim that familiarity with this task can lead to its gradual mastery, and eventually, to a higher level performance. The item concerning perceived task difficulty showed that these participants believed it was not too demanding for their level. As to the perception of the individual setting, the mean attitudinal score was slightly lower than that of the collaborative mode. However, despite the difficulties that learners in Comp experienced when reflecting in writing on linguistic issues during the reconstruction stage (as discussed above, in RQ1), these participants had a positive opinion of the written languaging task. At the same time, FFI+Collab expressed a favorable opinion of the pretask FFI they carried out before the DG. Nonetheless, the results pertaining to learners' perceived task difficulty and their opinion about wLRE and pretask FFI should be interpreted with caution, since they were only targeted by a single questionnaire item.

As in the case of the AQ (discussed above), we also found an advantage of Collab over the other two conditions. In fact, Collab obtained significantly higher scores than the other two groups with regards to DG task preferences, and they were also significantly more satisfied with the collaborative setting than FFI+Collab. If this finding is interpreted in parallel with learners' dyadic patterns (discussed above, in RQ2a), it does not come as a surprise that it was, in fact, FFI+Collab that displayed less collaborative engagement and more instances of conflicts when reconstructing the original texts. The examples of low mutuality and equality interactions during DG could have negatively impacted some of these learners' perception of the collaborative mode.

Learners' open-ended answers in the DQ and their responses to the focus group questions served to further confirm many of these findings. For instance, pair work ranked as the most preferred aspect of the DG task for those learners in Collab and FFI+Collab, in line with Kopinska and Azkarai (2020). Conversely, learners in Comp showed the highest preference for the audio recordings and the listening stage. It was noteworthy too that learners' favorite task features corresponded to their perceived task goals. Hence, learners who had experienced DG in collaboration most often mentioned that the task probably aimed to develop their collaborative skills (i.e. a non-linguistic goal, R. Ellis, 2003), while those in Comp claimed that it targeted their listening comprehension. Nonetheless, developing their writing expertise was not perceived as a major DG goal, and furthering grammar knowledge was only mentioned by a few learners in FFI+Collab. Moreover, echoing again the findings by Shak (2006) and Kopinska and Azkarai (2020), these learners most often referred to the writing stage as their least preferred part. In fact, their dislike towards the text reconstruction stage, be it individual or in pairs, can be understood inasmuch as they already expressed negative emotions towards L2 writing in the AQ. In the case of FFI+Collab, pair work was also most cited among the least preferred parts, which underpins the fact that, in this specific setting, the lower collaborative engagement shown during DG negatively affected their perception of pair work.

Nevertheless, the qualitative data sometimes offered certain nuances which were difficult to obtain by means of the questionnaire items. For example, albeit learners' favorable view on the benefits of TR expressed in the DQ, during the focus groups, there were some remarks about having a second DG enactment. In fact, some participants complained that it was too repetitive and unmotivating to carry out the task in the exact same condition. Thus, they suggested alternating peers in the dyads or even shifting from a collaborative to an individual working mode. Likewise, with regards to perceived task difficulty, the positive results from the DQ were also mitigated by some learners' comments during the interviews. Thus, most of their negative observations were related to the difficulties in understanding the audio recording and in recalling the ideas of the original texts. Finally, some child learners praised the whole experimental procedure, as they perceived it as way to experience the L2 in a different manner, a break from the school routines and also a fun activity (García Mayo, 2021a).

RQ3c What is the impact of L1 writing skills and attitudes towards writing on L2 writing?

In the current study, as discussed above in RQ2, we could not find enough evidence in either the text-based or rubric measures to claim that the DG task had an impact on these YLs' L2 writing quality (with the exception of some accuracy gains in FFI+Collab). However, in the analysis of learners' individual written production, it could also be observed that in some cases learners' output differed significantly right before the DG intervention, and what is more, that these differences were kept constant across the three time points (pretest, posttest and delayed posttest). For instance, Collab obtained significantly higher scores in adequacy, coherence, mechanics and lexis as compared to FFI+Collab. Why YLs' with a similar age, L1 background, L2 exposure, instruction type and proficiency could differ significantly in that many writing dimensions is a question we explored in the last research question. In other words, we intended to shed light on how two IDs, more specifically, learners' L1 writing quality and their attitude towards writing (discussed above, in RQ3a) and, could help explain the individual variation in L2 writing scores (in terms of lexical diversity), which was present from the onset of the experimental procedure.

In order to determine the influence of these two predictor variables (i.e. attitudes towards writing and L1 writing quality), a multiple regression analysis was conducted. In our model, child learners' attitudes towards writing explained more of the variance of L2 writing quality scores than their L1 writing proficiency. Although previous research in the topic is scant, the little contribution of L1 writing to the model (accounting for only 6% of all the variance) may be interpreted in the light of the Linguistic Threshold Hypothesis (Alderson & Urquhart, 1984; Clarke, 1980; Yamashita, 2001). This hypothesis moderated Cummins' original theory of Linguistic Interdependence (Cummins, 1979) by arguing that, for an efficient transfer of L1 skills to the L2 to happen, learners should have first reached a certain proficiency level in the L2. Therefore, as Pae (2018) reported in his study on the impact of L1 writing skills on Korean adolescent EFL learners' L2 writing, it could be the case that the low L2 proficiency of the participants (in the present study, falling under A1) was hindering a transfer of their L1 writing skills to the L2. However, this low correlation between L1 and L2 writing scores does not agree with Schoonen et al.'s (2011) longitudinal findings from the Dutch secondary school context. In fact, the authors reported that it was in the first years of this education stage (when learners were 13-14 years) that there was a stronger correlation between L1 writing in Dutch and EFL writing.

As Schoonen et al. (2011) explain, "these differences in outcomes, especially with respect to the L1 and FL writing correlation, are most likely caused by different operationalizations of writing proficiency, different languages and populations of writers" (p. 38). Contrary to the two aforementioned studies (Pae, 2018; Schoonen et al., 2011), who used holistic scores for rating the written output of learners in English, in the current study a lexical diversity index was employed, more specifically, Guiraud's index (Guiraud, 1959). This choice was motivated by previous research that had found lexical diversity to be a strong predictor of EFL writing quality (Crossley & McNamara, 2012; Treffers-Daller et al., 2016). However, L2 vocabulary may be associated with more language-specific knowledge than other higher-order, less language-specific knowledge indicators, such as the ability to create adequate or coherent texts (Cummins, 2008). Therefore, it could occur that while a holistic rating may allow to capture the result of higher-order ability writing knowledge, Guiraud's index for lexical diversity may be less permeable to L1 transfer (i.e. reflect language-specific knowledge to a greater extent).

Regarding the impact of attitudinal variables on the quality of L2 writing, our results suggest that affective variables play a role in predicting the EFL writing quality of YLs. Although the analysis of affective variables in relation to L2 writing achievement has not been widely studied in the past, and even less so in the early EFL context, our findings concur with those of recent research. For example, Pae (2008) found evidence for some indirect impact of attitudes towards writing on adult Korean EFL learners' writing. Moreover, Lee (2020), in a Korean adolescent EFL setting, reported that adding affective variables (such as cognitive/linguistic value and writing apprehension) to linguistic predictors (such as learners' reading ability and grammar knowledge) significantly increased the explained variance of L2 writing scores in their statistical model.

To conclude, the significant contribution of attitudes to L2 writing achievement underscores the need to better understand YLs' affective dimension in the EFL classroom (Li et al., 2018). In fact, it has been argued that teaching practices, including the type of tasks children deal with, influence their motivation to a greater extent than in the case of adult learners, who can find sources of motivation outside the classroom (Mihaljević Djigunović & Nikolov, 2019). Therefore, our finding contributes to the still scarce body of research that has examined affective factors as one of the many IDs involved in the variation of linguistic achievements in early FL programs.

CHAPTER 8 CONCLUSION: IMPLICATIONS, LIMITATIONS AND FUTURE RESEARCH

In this final chapter, we summarize the aims of the present dissertation, as well as the main research findings derived from the data analysis and the discussion. As one of the objectives of SLA research is to inform L2 pedagogy, some implications for early EFL practitioners will be outlined. In addition, we will acknowledge certain limitations of the current study and provide further lines for future research.

8.1 Conclusions

Although studies on CW are on the increase, there is still a paucity of research regarding early L2 learning settings and also with regards to the influence of learner internal and learner external variables. Therefore, the present study set out to shed light on these underexplored topics by investigating the potential of the DG task in the early EFL classroom. More specifically, we aimed to determine its efficacy in attracting YLs' (ages 11-12) attention to form, improving their writing quality and target form accuracy. The study examined the DG task potential in relation to two task-related variables: TR and the provision of pretask FFI. Besides, we aimed to examine several IDs, including YLs' initial disposition for collaborating and writing, their views on the task and the influence of L1 writing quality on predicting L2 writing achievement.

The analysis of the explicit languaging in the form of written and oral LREs served as evidence for learners' attention to form during the DG task. On the one hand, learners who performed the task individually (Comp) struggled to reflect their metalinguistic reflections in writing, but when they did so, most of their comments were related to lexis and mechanics. On the other hand, learners who experienced the task in the collaborative condition, generated most LREs about mechanics and other grammatical forms which were not targeted by the DG texts. Conversely, if the number of turns is taken into consideration, lexical LREs were the lengthiest ones, followed by F-others.

With regards to the influence of TR, repeating the task had in general a low impact on learners' amount of attention to form. However, it did lead to a significant decrease of the time spent on the task in one of the experimental conditions (FFI+Collab). As far as the pretask FFI is concerned, we failed to find a strong influence of this stage on YLs' languaging. In fact, although learners in FFI+Collab spent significantly more time on the task and attended significantly more to lexis than their counterparts in Collab, these two findings are difficult to relate exclusively to the influence of the FFI stage. Instead, they could be linked to other confounding variables, such as the setting (which resembled more a real classroom environment than in the case of Collab) or learners' prior vocabulary knowledge (which may have been greater in Collab, as attested by their proficiency scores).

Nonetheless, the pretask FFI did appear to influence these child learners' attention towards the target features at least in the case of 3S, as they produced on average significantly more episodes on this target form than Collab. Three different reasons could explain why providing learners with FFI seemed to be more beneficial in the case of the 3S than POSS: (i) there were more 3S than POSS target form instances in the DG texts, (ii) learners were likely to have received more prior instruction on 3S than POSS, as well as more exposure to it, and (iii) the context of POSS in the original story (i.e. kinship relationships) was relatively difficult, as EFL learners tend to acquire it in later developmental stages.

In addition, regardless of TR and FFI, YLs in this study correctly resolved the vast majority of their oral discussions, while the average of incorrectly resolved and unresolved LREs remained low. Regarding engagement, child participants generated more simple than elaborate episodes, with the exception of lexical LREs. It was also in Lex that these learners made an extensive use of their PKL. In fact, learners tended to favor the use of PKL in focus categories other than grammar. All in all, these results could imply that learners showed more engagement when they discussed linguistic issues in their PKL than in the L2, and that they probably lacked the necessary resources to discuss grammar in depth both in their PKL and L2.

The second topic we intended to shed some light on was the impact of collaboration on YLs' L2 writing. Firstly, the dyadic interactions on both DG days were qualitatively assessed. Our findings indicated that most of the learners in Collab and FFI+Collab displayed collaborative patterns. Repeating the task influenced these behaviors minimally, and in some cases, the shifts did not imply a greater level of mutuality and equality on Day 2. It was also interesting to note that learners in FFI+Collab showed less collaborative engagement than those in Collab.

Secondly, when the text reconstructions were analyzed, unexpectedly, we could not find an advantage of the collaborative setting (Collab and FFI+Collab) over the individual mode (Comp) in any of the text-based or rubric measures. Yet, TR positively influenced grammar accuracy in Collab, and especially, in FFI+Collab. In the second DG enactment, all child learners, regardless of their experimental condition, produced significantly less coordination, which was interpreted as a sign of a more adequate punctuation and segmentation. As far as target form accuracy is concerned, FFI+Collab obtained the highest accuracy rate, although it failed to reach statistical significance. In fact, based on the three aforementioned aspects (higher grammatical accuracy from Day 1 to Day 2 in Collab and FFI+Collab, fewer coordinated clauses in general and higher target form accuracy in FFI+Collab), we could tentatively suggest an impact of mechanical and grammar-focused LREs on the quality of the written production.

The third research question of this block was related to the effect of the DG task on the individual L2 writing quality. In general, we could not report a strong influence of the DG intervention, with the exception of grammar accuracy in FFI+Collab. In fact, these YLs produced significantly more grammatically accurate and contextualized clauses after the DG procedure, and furthermore, they continued to increase this rate in the delayed posttest. In this vein, although the analysis failed to show a significant difference, this pattern was endorsed by other text-based and rubric measures, as the participants in FFI+Collab also generated fewer grammar errors, more grammar error free clauses and higher target form accuracy rate after the two DG enactments.

Finally, in the present study, some IDs were also examined. Before the experimental procedure, child learners displayed a positive initial disposition towards collaboration and writing, although they were also less inclined to write in English than in Spanish. What is more, learners in FFI+Collab showed a less favorable attitude towards writing than the rest of the groups. Regarding the DG task, these YLs had in general a positive opinion about its main features (listening, note-taking and reconstructing), TR and perceived task difficulty. Nevertheless, learners in the individual setting (Comp) were slightly less enthusiastic about their own condition than those in the collaborative setting. Among the latter, Collab had a better opinion of their own setting than FFI+Collab, which goes in line with the less collaborative engagement shown by this group during DG. It could also be observed that learners in the collaborative setting appreciated the opportunities for pair work offered during the task, whilst those in Comp liked the most passive stage (i.e. the listening stage). In contrast to this, the text reconstruction stage ranked as one of the least preferred in all three conditions. In the open-ended questions and focus group interviews, learners confirmed many of these opinions, but they also expressed some remarks, including the fact that repeating the task in the same pair was perceived as demotivating and that the audio recording was difficult to understand.

To conclude, in an attempt to account for individual variation, we also examined the extent to which learners' L1 writing quality and attitudes towards writing could explain their L2 writing scores. The statistical model showed that child learners' attitudes explained a larger proportion of the variance than their L1 writing quality, highlighting the importance of the affective dimension in the early EFL classroom.

8.2 Pedagogical implications

Despite being widely used as a research tool to gather interactional data in CW research (Storch, 2019a; Zhang & Plonsky, 2020), the DG task was originally conceived as a pedagogical resource for EFL and ESL teachers (Wajnryb, 1990), whose main goals were to help learners use their productive grammar in text creation, encourage their understanding of what they know and do not know in English, and refine their use of the written language.

After other scholars empirically examined the applicability of the DG task with adult and adolescent EFL learners (as thoroughly reviewed in the literature section), the present dissertation adds to the recent cumulative evidence (Calzada & García Mayo, 2020a, 2020b, 2021a, 2021b, 2021c) showing that it may also be appropriate for low proficiency EFL learners as young as 11-12 years. In fact, across several studies, these learners have shown that they focus primarily on mechanics and grammar, and are able to correctly resolve most of their discussions. Therefore, as a general recommendation, teachers may wish to employ the DG task for developing children's grammar skills, but also for improving their awareness of spelling and punctuation. Furthermore, although mechanical aspects were not part of the original task goals, recently some researchers have explored this path by slightly adapting the DG procedure (turning it into task known as the "zero-error dictation" or metacognitive dictation) with young L1 French learners in Canada (aged 8-12) (Ammar & Hassan, 2018; Nadeau, Arseneau, et al., 2020; Nadeau, Quevillon Lacasse, et al., 2020).

We will now interpret the main findings from the current study in terms of their applicability in the early EFL classroom and provide guidelines for practitioners. For DG to succeed in diverting YLs' attention to accuracy, it is necessary that certain aspects of its design are taken into consideration. First and foremost, learners should be familiar with the topic and the vocabulary included in the passages. In the current study, the lexical items that were key for the correct understanding of the passage were mostly related to family relationships and school routines.

Moreover, it is also recommendable that child learners are provided with some vocabulary help before they start reconstructing the text. For instance, in Calzada and García Mayo (2020a), akin to Wajnryb's (1990) original procedure, we introduced a warm-up stage in which the participants were shown some pictures related to the DG topics, and were asked questions in order to elicit that vocabulary from them and make them more personally involved. In the current dissertation, however, a picture summarizing the main characters and plot of the DG passage was included in the note-taking photocopy. This aid probably led learners to free more attentional resources so that they could focus on language. Yet, in the case of FFI+Collab, as discussed above, this support may not have been sufficient, since they produced numerous lexical episodes and expressed difficulties in understanding parts of the story. Finally, teachers should also consider recycling short texts that learners have already worked on for different purposes, such as reading comprehension activities. By doing so, learners will be more likely to focus on formal aspects of language during the text reconstruction.

Furthermore, it should also be pointed that, albeit learners' primary focus on grammar and mechanics, most of their discussions reflected a shallow engagement, which was also to a certain extent associated with a greater use of the L2 for languaging. This finding raises two questions which may be of interest for practitioners: on the one hand, to what extent YLs should be allowed to use the L1, and on the other, what possible ways there are to increase children's engagement.

Regarding L1 use, in the present study child participants made an extensive use of their PKL to discuss topics other than grammar, such as lexis, mechanics and discourse. Instructors who are to implement DG in the low proficiency EFL classroom should remain flexible about their learners' use of the L1 while collaborating to write a text in English. As previous research has shown (Payant, 2020), multilingual writers fall back on their whole linguistic repertoire when constructing a text and, therefore, the use of PKL should be perceived as a cognitive tool (Martínez Adrián & Arratibel Irazusta, 2020), in line with the TBLT precepts arguing that learners should draw on their existing linguistic resources (both the L1 and the L2) for comprehension and production (R. Ellis et al., 2020). Moreover, the written modality also plays a role, as learners might consider that they are already producing the L2 in their writing and, hence, allow themselves a greater use of the L1 when interacting orally.

Nevertheless, YLs with a low proficiency may sometimes lack the necessary resources to discuss certain linguistic issues in depth, resulting in short turns and brief episodes. Although impoverished language use has been one of the criticisms made about TBLT, R. Ellis (2009), for instance, argued that, when learners are beginners such interaction might be beneficial, as it can encourage their capacity to make use of limited resources and help them develop their strategic competence. Similarly, other researchers have explored the possibility of providing students with tools to become better learners in collaborative tasks (Fujii et al., 2016). These resources have usually been labelled "metacognitive instruction", and Sato has been one of the researchers that has examined the application of this instruction in different EFL contexts, including Chilean EFL adult (Sato & Loewen, 2018) and adolescent learners (Sato, 2020a). More recently, Sato and Dussuel Lam (2021) investigated the use of metacognitive instruction with young EFL learners in Chile (aged 8-9) who performed an oral communicative task. The instruction was aimed at increasing their use of the L2, self-monitoring and developing conversation strategies. Their findings showed that the group that received metacognitive instruction produced more turns in the L2 and managed their conversation in a more efficient way.

Likewise, learners' ability to run their collaborative dialogue smoothly is linked to the type of dyadic patterns, and more specifically, to the level of mutuality and equality displayed during the task. In our study, unlike in previous work by Azkarai and Kopinska (2020), most patterns fell under the collaborative pattern, although there were also several examples of cooperative (or passive/parallel) interaction, especially in the FFI+Collab group. Moreover, there were also instances of low equality between learners in the same dyad, which were usually described as dominant/dominant or dominant/passive, and were characterized by frequent arguments and off-task conversations. When pairs do not work efficiently, an easy solution can be changing partners. However, if the same children continue to trigger uncollaborative behavior, the problem may lie in their understanding of the value of the DG task. In this respect, metacognitive instruction has been proposed as a possible intervention. For instance, Chen and Hapgood (2019) applied a series of metacognitive activities to raise adult ESL learners' awareness of the benefits of a CW task. As a result, learners who received this instruction displayed more collaborative patterns and generated more LREs. The extent to which these findings may be extrapolated to early EFL settings need still to be studied, but given the positive results of metacognitive instruction with children performing oral tasks (Sato & Dussuel Lam, 2021), they offer a promising avenue for practitioners.

With regards to use of a pretask FFI stage before the DG task we found that FFI+Collab participants produced more episodes about the target forms and that they showed a higher written accuracy. Moreover, they also reflected an upward trend in general grammar accuracy measures from the pretest to the posttest. Yet, their advantage in languaging over Collab was only significant in the case of 3S, but not in POSS. Thus, we could tentatively suggest that FFI prior to DG may lead to benefits provided that there are enough target instances in the DG text, that learners are already familiar with the target features and that they are developmentally ready to master that target form (requirements that POSS did not fully meet in our case). In the present study, learners were not allowed to keep their FFI worksheets during the text reconstruction, but there is evidence from an individual DG investigation indicating that pretask FFI combined with during task FFI leads to greater accuracy than pretask FFI only (Khezrlou, 2019). Hence, teachers may also want to explore providing children with greater FFI support during collaborative DG.

As far as TR is concerned, participants in our study had a positive opinion of having a second attempt at the DG task. Although this favorable belief did not translate into a better quality writing, coordination, which learners tended to overuse in their texts, was less present on the second day. Furthermore, the combination of collaboration and TR seemed to lead to increased grammatical accuracy in the second enactment. Apart from the learners' inclination to repeat the task and some indication of grammar improvement, there is more justification for implementing DG in the classroom on a regular basis. When we compared the type of clauses learners produced in their reconstruction and in their individual stories (as part of the pretest, posttest and delayed posttest), we observed that the proportion of full-fledged clauses was much greater in the former than in the latter. In other words, having a predetermined content (i.e. dealing with a content-reproducing task) led these learners to produce better texts. Consequently, the repeated practice of successful L2 writing experiences could help child learners gain confidence in their English writing skills and lessen their anxiety (Schunk, 1984). In fact, triggering positive emotions towards L2 writing has been demonstrated as being very relevant in our study, as these attitudes explain part of their L2 writing quality.

However, TR may also lead to higher levels of demotivation to carry out the task. Therefore, teachers should strike a balance, for instance, by alternating individual and collaborative DG (Kopinska & Azkarai, 2020). In fact, learners have also shown appreciation for individual writing, as it promotes concentration and it allows them to make their own decisions. In this sense, if individual DG is to be implemented together with a written languaging stage, child learners need sufficient prior training, as our study and other authors (Luquin & García Mayo, 2021) have observed that learners aged 11-12 struggle to put in writing their metalinguistic reflections.

In summary, DG appears to suit young EFL learners in the last years of Primary Education inasmuch as it encourages their attention to form and accuracy, and it matches these learners' favorable disposition towards peer collaboration. We hope that the guidelines discussed above will orient teachers' decisions when implementing this task in their classroom and that they will contribute to improving L2 grammar and writing instruction in early EFL settings.

8.3 Limitations and further lines of research

The present dissertation aimed to shed light on the under-explored area of collaborative L2 writing in early EFL settings. However, there are a number of limitations to this research which should be considered when interpreting its findings. First, we only used one type of collaborative task, namely, the DG task, with a very specific population (young EFL learners aged 11-12 in the BAC). Hence, the outcomes of the study and their generalizability are restricted to the kind of task we used, and to the population of learners involved. For instance, it is likely that if the DG is performed by younger learners the results may vary, as they will be at a different stage of cognitive and behavioral development (Pinter, 2011).

Regarding the three task conditions (Comp, Collab and FFI+Collab), the numerous difficulties in coordinating our schedule with children's regular teaching led to a quasi-experimental design, which implied that learners' DG procedure was determined by the school that they attended. Ideally, future studies should overcome this shortcoming by using an experimental approach (in which participants are assigned to different conditions randomly), or even better, by means of a within-participants design, whereby all learners experience all conditions. In this way, the effect of condition is more clearly observed, since in between-participants designs there is a risk that extraneous variables (such as the teaching practices of each school or classroom) influence the results.

The current study lacked a true control group that did not experience the DG task and, therefore, when analyzing the task influence on the development of L2 writing, our claims are restricted to the comparison of learners experiencing the task individually and those performing it collaboratively. Furthermore, the comparison of the reconstructed texts was negatively affected by having unbalanced groups. Although the number of participants was similar, the analysis took into consideration the written products and in the collaborative conditions (Collab and FFI+Collab), the sample was half the one corresponding to the individual setting. In this sense, Elabdali (2021) considers this unbalance one of the major drawbacks in studies examining collaborative versus individual writing modes, and suggests that the sample of pairs should be double to equal that of the individual condition. Yet, in the BAC context, schools in the capital cities usually have three classes per academic year, and there cannot be more than 25 children in the classroom (Order of 27th April 2016). Consequently, in our case, it would have implied having a third school of similar characteristics in order to recruit a sufficient number of participants. As we specified in the procedure section, research with child participants entails a laborious process for obtaining the necessary permission from the Ethical Committee, the school board and the parents, which is obviously more complex the more schools that are involved. In fact, we already feel grateful for being granted access to the two schools that took part in the current study.

The last limitation related to the task conditions is the fact that the setting in FFI+Collab resembled more that of a real classroom than Collab's, as in the former, five pairs performed the DG task simultaneously, whereas in the latter pairs were withdrawn separately to a different classroom with the only presence of the researcher or one of the research assistants. Laboratory settings are considered to help researchers guarantee good recording conditions for collaborative tasks (Collins & White, 2019). However, we also noticed that pairs in Collab were sometimes too shy to speak their minds because of the silent setting, whereas children in FFI+Collab felt more at ease and seldom did the researcher ask them to raise their voice. Although we would have wished for both experimental groups to have more similar conditions, the need to concentrate the same number of sessions in the same time period relying on the same human resources did not allow a greater homogenization, as the FFI stage would have required extra time per pair. In this vein, we would also like to make a claim for more classroom-based studies, as they would serve to validate our findings.

With regard to the units of analysis, the operationalization of languaging in the form LREs, a concept widely employed in collaborative tasks with adolescents and adults, proved to be appropriate for examining children's collaborative dialogue. Furthermore, its popularity in SCT informed research allows for the comparability of results across studies. In contrast, a problem related to the study of learners' L1 (or PKL) use during collaborative tasks, as discussed above, is the various measures previous researchers have employed, including predominance of words in one of the languages (the criterium used in the current study) or a clause-based analysis. Choosing one or the other can lead to different results compromising comparability.

In the case of text-based complexity and accuracy measures, we employed traditional measures (such as the number of clauses per T-unit or errors per 100 words). Yet, these measures, despite indicating certain trends (e.g. grammatical accuracy improvement in FFI+Collab), in general failed to capture any significant change in children's writing. In this sense, some authors who have also looked into the quality of child EFL learners' writing have suggested using alternative measures of writing for this population, such as rubrics for global assessment (Hidalgo & Lázaro Ibarrola, 2020). Nonetheless, our results indicate that text-based measures should not be completely dismissed, but rather adapted for child low proficiency learners' writing. For instance, we opted for including a measure of coordination for complexity, which has been recommended when assessing beginner learners' production (Bardovi-Harlig, 1992; Bulté & Housen, 2012; Mylläri, 2020). In fact, coordination allowed us to observe a significant decrease in children's reconstructions from Day 1 to Day 2. Moreover, the analysis of the type of clauses following Torras' (2005) classification also led to some interesting comparison between children's performance in a content-reproducing and a content-generating task.

In addition to focusing on the written product, it has been argued that researchers should also examine the writing processes involved (Michel et al., 2020; Révész, 2019; Révész et al., 2017). In the case of CW, learners' collaborative dialogue enabled us to observe part of that planning, language monitoring and revising process, whereas in the case of individual writing it is more complicated, and different techniques have been proposed, including keystroke logging, eye-tracking or introspective methods, such as stimulated recall or think-aloud protocols. Obtaining a better insight of learners' behaviors during writing tasks can, indeed, inform teachers' pedagogical decisions. For instance, a child learner who struggles with vocabulary may need a different kind of instruction from another who has difficulties with grammar.

As far as the instruments to tap into learners' attitudes are concerned, there are two main points to be discussed about the questionnaires. On the one hand, the AQ left out of the analysis the dimension of "English learning" due to a low internal subscale reliability, associated with the inclusion of five subconstructs that were targeted by only eight items. However, this does not imply that child learners' attitudes towards L2 instruction are not influencing their L2 writing. Future studies may want to include this variable by increasing the number of items or reducing the number of subconstructs. On the other hand, some constructs from the DQ, such as learners' perception of task difficulty or their opinion about the FFI stage, were only targeted with a single item. Hence, in further investigations, more varied measured should be used. Despite these limitations, both the questionnaires and the focus group interviews constituted very appropriate instruments for EFL learners aged 11-12, as they were able to complete the reports autonomously and were eager to tell their opinion in the interviews. Yet, it should be reminded that AQ and DQ are far from being universal, and if researchers wish to apply them in different teaching contexts some adaptation may be needed.

In addition, regarding the target linguistic features (3S and POSS), our study only assessed their development by means of written accuracy in the story continuation tasks and the text reconstructions. It would be interesting to analyze if the DG procedure had any impact in their implicit and explicit knowledge of the target forms (Khezrlou, 2021). Additionally, in order to further explore the influence of TR, we suggest increasing the number of repetitions, or eventually, taking a longitudinal approach to the study of CW, as examples of this kind are still scant in the literature (Shehadeh, 2011). An area that also clearly deserves more attention is the relationship between learners' affective dimension and their propensity to initiate metalinguistic discussions during collaborative work, as there is evidence showing that learner attitudes influence their interactional behavior (Sato, 2020b; Sato & McDonough, 2020). Similarly, an exploration of pair dynamics in relation to attention for form (Azkarai & Kopinska, 2020) and the text quality (Zhang, 2019b) is recommended.

Despite these limitations, we would like to consider the study reported in the present dissertation a step forward in the area of CW and a contribution to inform both the emerging research on early EFL settings, as well as teachers' pedagogical decisions. Needless to say that the resultant proposals would clearly benefit from future investigations.

REFERENCES

Abrahamsson, N., & Hyltenstam, K. (2008). The robustness of aptitude effects in near-native second language acquisition. *Studies in Second Language Acquisition*, *30*(4), 481–509. https://doi.org/10.1017/S027226310808073X

Agustín Llach, M. P. (2011). *Lexical errors and accuracy in foreign language writing*. Multilingual Matters. http://site.ebrary.com/id/10492855

Ahmadian, M. J., & Mansouri, A. (2020). Using L1 for collaborative pre-task planning. In C. Lambert & R. Oliver (Eds.), *Using tasks in diverse contexts* (pp. 237-254). Multilingual Matters.

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

Ahmadian, M. J., Mansouri, S. A., & Ghominejad, S. (2017). Language learners' and teachers' perceptions of task repetition. *ELT Journal*, *71*(4), 467–477. https://doi.org/10.1093/elt/ccx011

Aitchison, J. (2003). *Words in the mind: An introduction to the mental lexicon* (3rd ed). Blackwell Pub.

Ajzen, I. (1988). *Attitudes, personality and behavior*. Open University Press.

Alderson, J. C., & Urquhart, A. H. (1984). *Reading in a foreign language*. Longman Pub Group.

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

Amiryousefi, M. (2016). The differential effects of two types of task repetition on the complexity, accuracy, and fluency in computer-mediated L2 written production: A focus on computer anxiety. *Computer Assisted Language Learning*, *29*(5), 1052–1068. https://doi.org/10.1080/09588221.2016.1170040

Ammar, A., & Hassan, R. M. (2018). Talking it through: Collaborative dialogue and second language learning. *Language Learning*, *68*(1), 46–82. https://doi.org/10.1111/lang.12254

Anam, S., & Stracke, E. (2016). Language learning strategies of Indonesian primary school students: In relation to self-efficacy beliefs. *System*, *60*, 1–10. https://doi.org/10.1016/j.system.2016.05.001

Apraiz Jaio, M. V., Pérez Gómez, M., & Ruiz Pérez, T. (2012). La enseñanza integrada de las lenguas en la escuela plurilingüe. *Revista Iberoamericana de Educación*, *59*, 119–137. https://doi.org/10.35362/rie590459

Artiles, C., & Jiménez, J. E. (2007). *Proescri Primaria. Prueba de evaluación de procesos cognitivos en la escritura*. Canarias: Consejería de Educación, Cultura y Deportes del Gobierno de Canarias. Dirección General de Ordenación e Innovación Educativa.

Ashcraft, N. (2018). Information-Gap Tasks. In J. I. Liontas, T. International Association, & M. DelliCarpini (Eds.), *The TESOL Encyclopedia of English Language Teaching* (pp. 1–5). John Wiley & Sons, Inc. https://doi.org/10.1002/9781118784235.eelt0190

Aula Blasco, J. (2016). The relationship between writing anxiety, writing self-efficacy, and Spanish EFL students' use of metacognitive writing strategies: A case study. *Journal of English Studies*, *14*(0), 7–45. https://doi.org/10.18172/jes.3069

Azkarai, A., & García Mayo, M. P. (2015). Task-modality and L1 use in EFL oral interaction. *Language Teaching Research*, *19*(5), 550–571. https://doi.org/10.1177/1362168814541717

Azkarai, A., & García Mayo, M. P. (2017). Task repetition effects on L1 use in EFL child task-based interaction. *Language Teaching Research*, *21*(4), 480–495. https://doi.org/10.1177/1362168816654169

Azkarai, A., García Mayo, M. P., & Oliver, R. (2019). The effect of task repetition on the patterns of interaction of ESL children. *ITL - International Journal of Applied Linguistics*. https://doi.org/10.1075/itl.17034.azk

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

Azkarai, A., & Oliver, R. (2018). Teaching EFL to Young Learners. In R. Oliver & A. Azkarai (Eds.), *Teaching young second language learners* (pp. 89–108). Routledge.

Baleghizadeh, S., & Farhesh, S. (2014). The impact of pair work on EFL learners' motivation. *MEXTESOL Journal*, *38*(3), 1–11.

Baralt, M. (2014). Task complexity and task sequencing in traditional versus online language classes. In M. Baralt, R. Gilabert, & P. J. Robinson (Eds.), *Task sequencing and instructed second language learning* (pp. 97–122). Bloomsbury Academic.

Baralt, M., Gurzynski-Weiss, L., & Kim, Y. J. (2016). Engagement with the language: How examining learners' affective and social engagement explains successful learner-generated attention to form. In M. Sato & S. Ballinger (Eds.), *Language Learning & Language Teaching* (Vol. 45, pp. 209–239). John Benjamins Publishing Company. https://doi.org/10.1075/lllt.45.09bar

Bardovi-Harlig, K. (1992). A second look at T-Unit analysis: reconsidering the sentence. *TESOL Quarterly*, *26*(2), 390. https://doi.org/10.2307/3587016

Barnes, D. (1976). *From communication to curriculum*. Penguin.

Basterrechea, M., & Gallardo-del-Puerto, F. (2020). Language-related episodes and pair dynamics in primary school CLIL learners: A comparison between proficiency-matched and student-selected pairs. *Studies in Second Language Learning and Teaching*, *10*(3), 423–447. https://doi.org/10.14746/ssllt.2020.10.3.2

Basterrechea, M., & García Mayo, M. P. (2013). Language-related episodes during collaborative tasks: A comparison of CLIL and EFL learners. In K. McDonough & A. Mackey (Eds.), *Second Language Interaction in Diverse Educational Contexts* (pp. 25–43). John Benjamins Publishing Company. http://www.jbe-platform.com/content/books/9789027272348-05ch2

Bialystok, E. (2001). *Bilingualism in development: language, literacy, and cognition* (1st ed.). Cambridge University Press. https://doi.org/10.1017/CBO9780511605963

Biber, D., Gray, B., & Poonpon, K. (2011). Should we use characteristics of conversation to measure grammatical complexity in L2 writing development? *TESOL Quarterly*, *45*(1), 5–35. https://doi.org/10.5054/tq.2011.244483

Bitchener, J. (2008). Evidence in support of written corrective feedback. *Journal of Second Language Writing*, *17*(2), 102–118. https://doi.org/10.1016/j.jslw.2007.11.004

Bitchener, J. (2016). *Why written corrective feedback can contribute to L2 development: A theoretical model*. 13613 Bytes. https://doi.org/10.4225/03/58350C08EC7D1

Bitchener, J., & Storch, N. (2016). *Written corrective feedback for L2 development*. Multilingual Matters.

Bland, J. (2019). Teaching English to young learners: More teacher education and more children's literature! *Children's Literature in English Language Education*, *7*, 79–103.

Bley-Vroman, R. (1988). The fundamental character of foreign language learning. In W. Rutherford & M. Sharwood Smith (Eds.), *Grammar and second language teaching: A book of readings* (pp. 19–30). Newbury House.

Bley-Vroman, R. (1989). What is the logical problem of foreign language learning? In S. Gass & J. Schachter (Eds.), *Linguistic perspectives on second language acquisition* (1st ed., pp. 41–68). Cambridge University Press. https://doi.org/10.1017/CBO9781139524544.005

Breen, M. (1989). The evaluation cycle for language learning tasks. In R. K. Johnson (Ed.), *The second language curriculum* (1st ed., pp. 187–206). Cambridge University Press. https://doi.org/10.1017/CBO9781139524520.014

Breen, M., & Candlin, C. N. (1980). The essentials of a communicative curriculum in language teaching. *Applied Linguistics*, *1*(2), 89–112.

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

Brown, Roger. (1973). *A first language: The early stages*. Harvard University Press Cambridge, Mass.

Bruner, J. S. (1973). *The relevance of education*. Norton.

Brunswick, E., Goldschneider, L., & Pilek, E. (1932). *Unterssuchungen zur Entwiklung des Gedachtnisses (Studies on the development of memory).* (W. Stern & O. Lippmann, Eds.). Ambrosius Barth.

Bryfonski, L., & McKay, T. H. (2019). TBLT implementation and evaluation: A meta-analysis. *Language Teaching Research*, *23*(5), 603–632. https://doi.org/10.1177/1362168817744389

Bueno-Alastuey, C., & Martínez de Lizarrondo Larumbe, P. (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.

Bulté, B., & Housen, A. (2012). Defining and operationalising L2 complexity. In *Dimensions of L2 performance and proficiency*. John Benjamins. https://www.jbe-platform.com/content/books/9789027273260-02bul

Bulté, B., & Housen, A. (2014). Conceptualizing and measuring short-term changes in L2 writing complexity. *Journal of Second Language Writing*, *26*, 42–65. https://doi.org/10.1016/j.jslw.2014.09.005

Butler, Y. G. (2016). 22. Assessing young learners. In D. Tsagari & J. Banerjee (Eds.), *Handbook of Second Language Assessment* (pp. 359-375). De Gruyter. https://doi.org/10.1515/9781614513827-024

Butler, Y. G. (2017). The role of affect in intraindividual variability in task performance for Young Learners. *TESOL Quarterly*, *51*(3), 728–737. https://doi.org/10.1002/tesq.385

Butler, Y. G. (2019). Assessment of young English learners in instructional settings. In X. Gao (Ed.), *Second Handbook of English Language Teaching* (pp. 477–496). Springer International Publishing. https://doi.org/10.1007/978-3-030-02899-2_24

Butler, Y. G., & Zeng, W. (2015). Young learners' interactional development in task-based paired-assessment in their first and foreign languages: A case of English learners in China. *Education 3-13*, *43*(3), 292–321. https://doi.org/10.1080/03004279.2013.813955

Bygate, M. (2009). Chapter 12. Effects of task repetition on the structure and control of oral language. In K. Van den Branden, M. Bygate, & J. M. Norris (Eds.), *Task-Based Language Teaching* (Vol. 1, pp. 249–274). John Benjamins Publishing Company. https://doi.org/10.1075/tblt.1.15eff

Bygate, M. (Ed.). (2015). *Domains and directions in the development of TBLT: A decade of plenaries from the international conference* (Vol. 8). John Benjamins Publishing Company. https://doi.org/10.1075/tblt.8

Bygate, M. (Ed.). (2018). *Learning language through task repetition*. John Benjamins Publishing Company.

Byrnes, H., & Manchón, R. M. (Eds.). (2014). *Task-based language learning – Insights from and for L2 Writing* (Vol. 7). John Benjamins Publishing Company. https://doi.org/10.1075/tblt.7

Calzada, A., & García Mayo, M. P. (2020a). Child EFL grammar learning through a collaborative writing task. In W. Suzuki & N. Storch (Eds.), *Languaging in language learning and teaching: A collection of empirical studies* (pp. 20–39). John Benjamins. https://doi.org/10.1075/lllt.55.01cal

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

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

Calzada, A., & García Mayo, M. P. (2021b). *Dictogloss to focus on possessive determiners in early EFL: Is it enough?* [Conference presentation]. 38th AESLA International Conference 2021, online. https://aesla2020.udc.es/

Calzada, A., & García Mayo, M. P. (2021c). Effects of proficiency and collaborative work on child EFL individual dictogloss writing. *Language Teaching for Young Learners*, *3*(2), 246-274. https://doi.org/10.1075/ltyl.20003.cal

Cambridge Assessment English. (n.d.-a). A2 Flyers results. https://www.cambridgeenglish.org/exams-and-tests/flyers/results/

Cambridge Assessment English. (n.d.-b). *The Cambridge young learners' programme*. https://cambridge-exams.ch/why-are-cambridge-english-young-learners-tests-ideal-children-primary-school

Cambridge Assessment English . (2018). *Young learners sample papers 2018—Flyers A2*. Cambridge Assessment English. https://www.cambridgeenglish.org/Images/young-learners-sample-papers-2018-vol1.pdf

Carless, D. (2002). Implementing task-based learning with young learners. *ELT Journal*, *56*(4), 389–396. https://doi.org/10.1093/elt/56.4.389

Carless, D. (2003). Factors in the implementation of task-based teaching in primary schools. *System*, *31*(4), 485–500. https://doi.org/10.1016/j.system.2003.03.002

Carless, D. (2004). Issues in teachers' reinterpretation of a task-based Innovation in primary schools. *TESOL Quarterly*, *38*(4), 639. https://doi.org/10.2307/3588283

Carrell, P. L. (1985). Facilitating ESL reading by teaching text structure. *TESOL Quarterly*, *19*(4), 727–752. JSTOR. https://doi.org/10.2307/3586673

Cenoz, J. (2012). Bilingual educational policy in higher education in the Basque Country. *Language, Culture and Curriculum*, *25*(1), 41–55. https://doi.org/10.1080/07908318.2011.653057

Cenoz, J., & Gorter, D. (2011). Focus on multilingualism: A study of trilingual writing. *The Modern Language Journal*, *95*(3), 356–369. https://doi.org/10.1111/j.1540-4781.2011.01206.x

Chen, Wenting, & Hapgood, S. (2019). Understanding knowledge, participation and learning in L2 collaborative writing: A metacognitive theory perspective. *Language Teaching Research*, 136216881983756. https://doi.org/10.1177/1362168819837560

Chen, Wenting, & Yu, S. (2019). A longitudinal case study of changes in students' attitudes, participation, and learning in collaborative writing. *System*, *82*, 83–96. https://doi.org/10.1016/j.system.2019.03.005

Chen, Wenxue. (2018). Patterns of pair interaction in communicative tasks: The transition process and effect on L2 teaching and learning. *ELT Journal*. https://doi.org/10.1093/elt/ccy015

Cho, Y., & So, Y. (2014). *Construct-irrelevant factors infl uencing young English as a foreign language (EFL) learners' perceptions of test task difficulty (Research Memorandum No. RM-14- 04).* Educational Testing Service. https://www.ets.org/s/research/pdf/RM-14-04.pdf

Choi, H., & Iwashita, N. (2016). Interactional behaviours of low-proficiency learners in small group work. In M. Sato & S. Ballinger (Eds.), *Language Learning & Language Teaching* (Vol. 45, pp. 113–134). John Benjamins Publishing Company. https://doi.org/10.1075/lllt.45.05cho

Chomsky, N. (1986). *Knowledge of Language: Its Nature, Origin, and Use*. Greenwood Publishing Group.

Clarke, M. A. (1980). The short circuit hypothesis of ESL reading-or when language competence interferes with reading performance. *The Modern Language Journal*, *64*(2), 203–209.

Collins, L., & White, J. (2019). Observing language-related episodes in intact classrooms: Context matters! In R. DeKeyser & G. Prieto Botana (Eds.), *Language Learning & Language Teaching* (Vol. 52, pp. 9–30). John Benjamins Publishing Company. https://doi.org/10.1075/lllt.52.02col

Collins, L., Trofimovich, P., White, J., Cardoso, W., & Horst, M. (2009). Some input on the easy/difficult grammar question: an empirical study. *The Modern Language Journal*, *93*(3), 336–353. https://doi.org/10.1111/j.1540-4781.2009.00894.x

Cook, V. (1992). Evidence for multicompetence. *Language Learning*, *42*(4), 557–591. https://doi.org/10.1111/j.1467-1770.1992.tb01044.x

Cook, V. (1998). Internal and external uses of a second language. *Essex Research Reports in Linguistics*, 100–110.

Cook, V. (2010). The relationship between first and second language acquisition revisited. In E. Macaro (Ed.), *The continuum companion to second language acquisition* (pp. 137–157). Continuum.

Costa, P., & Albergaria-Almeida, P. (2015). The European survey on language competences: measuring foreign language student proficiency. *Procedia - Social and Behavioral Sciences*, *191*, 2369–2373. https://doi.org/10.1016/j.sbspro.2015.04.255

Council of Europe. (2001). *Common European framework of reference for languages: Learning, teaching, assessment*. Press Syndicate of the University of Cambridge. https://rm.coe.int/1680459f97

Council of Europe. (2018). *Common European Framework of Reference for Languages: Learning, Teaching, Assessment Companion Volume with New Descriptors*. Council of Europe. https://rm.coe.int/cefr-companion-volume-with-new-descriptors-2018/1680787989

Coyle, Y., & Roca de Larios, J. (2014). Exploring the role played by error correction and models on children's reported noticing and output production in a L2 writing task. *Studies in Second Language Acquisition*, *36*(03), 451–485. https://doi.org/10.1017/S0272263113000612

Coyle, Y., Cánovas Guirao, J., & Roca de Larios, J. (2018). Identifying the trajectories of young EFL learners across multi-stage writing and feedback processing tasks with model texts. *Journal of Second Language Writing*, *42*, 25–43. https://doi.org/10.1016/j.jslw.2018.09.002

Crossley, S. A., & McNamara, D. S. (2012). Predicting second language writing proficiency: The roles of cohesion and linguistic sophistication. *Journal of Research in Reading*, *35*(2), 115–135. https://doi.org/10.1111/j.1467-9817.2010.01449.x

Csizér, K. (2017). Motivation in the L2 Classroom. In S. Loewen & M. Sato (Eds.), *The Routledge handbook of instructed second language acquisition* (pp. 418–432). Routledge. https://doi.org/10.4324/9781315676968.ch23

Cumming, A. (1990). Metalinguistic and ideational thinking in second language composing. *Written Communication*, *7*(4), 482–511. https://doi.org/10.1177/0741088390007004003

Cummins, J. (1979). Linguistic interdependence and the educational development of Bilingual Children. *Review of Educational Research*, *49*(2), 222. https://doi.org/10.2307/1169960

Cummins, J. (2008). Teaching for transfer: challenging the two solitudes assumption in bilingual education. In N. H. Hornberger (Ed.), *Encyclopedia of language and education* (pp. 1528–1538). Springer US. https://doi.org/10.1007/978-0-387-30424-3_116

Damon, W., & Phelps, E. (1989). Strategic uses of peer learning in children's education. In T. J. Berndt & G. W. Ladd (Eds.), *Peer relationships in child development.* (pp. 135–157). John Wiley & Sons.

Decree 236/2015, of 22nd December, by which the Basic Education Curriculum is established in the Basque Autonomous Community (2015). *Boletín Oficial del País Vasco,* 15th January 2016.

Decree 237/2015, of 22nd December, by which the Pre-primary Education Curriculum is established in the Basque Autonomous Community (2015). *Boletín Oficial del País Vasco,* 15th January 2016.

DeKeyser, R. (1998). Beyond focus on form: Cognitive perspectives on learning and practicing second language grammar. In C. Doughty & J. Williams (Eds.), *Focus on form in classroom second language acquisition*, (pp. 42–63). Cambridge University Press.

DeKeyser, R. (2000). The robustness of critical period effects in second language acquisition. *Studies in Second Language Acquisition*, *22*(4), 499–533. https://doi.org/10.1017/S0272263100004022

DeKeyser, R. (2003). Implicit and explicit learning. In C. J. Doughty & M. Long (Eds.), *The Handbook of Second Language Acquisition* (pp. 312–348). Blackwell Publishing Ltd. https://doi.org/10.1002/9780470756492.ch11

DeKeyser, R. (2007). *Practice in a second language: perspectives from applied linguistics and cognitive psychology*. Cambridge University Press. https://doi.org/10.1017/CBO9780511667275

DeKeyser, R. (2020). Input is not a panacea. *International Journal of Bilingualism, 24*(1), 79–81. https://doi.org/10.1177/1367006918768371

DeKeyser, R., & Larson-Hall, J. (2005). What Does the Critical Period Really Mean? In *Handbook of bilingualism: Psycholinguistic approaches* (pp. 88–108). Oxford University Press.

DeKeyser, R., Alfi-Shabtay, I., & Ravid, D. (2010). Cross-linguistic evidence for the nature of age effects in second language acquisition. *Applied Psycholinguistics*, *31*(3), 413–438. https://doi.org/10.1017/S0142716410000056

Department of Education. (2020). *ESE2 2019/2020 English literacy marking guidelines*. Government of Navarre. https://www.educacion.navarra.es/documents/27590/1678902/ESO2_Competencia_Ingles_19_20_CC.pdf/dcd33fbc-5b1e-4b28-478e-777128d30273

Dewey, J. (1997). *Experience and education*. Simon & Schuster.

Díez-Bedmar, M. B., & Pérez-Paredes, P. (2020). Noun phrase complexity in young Spanish EFL learners' writing: Complementing syntactic complexity indices with corpus-driven analyses. *International Journal of Corpus Linguistics*, *25*(1), 4–35. https://doi.org/10.1075/ijcl.17058.die

Doiz, A., Lasagabaster, D., & Sierra, J. M. (2014). CLIL and motivation: The effect of individual and contextual variables. *The Language Learning Journal*, *42*(2), 209–224. https://doi.org/10.1080/09571736.2014.889508

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

Donato, R. (2000). Sociocultural contributions to understanding the foreign and second language classroom. In J. P. Lantolf (Ed.), *Sociocultural theory and second language learning* (pp. 27-52). Oxford University Press.

Dörnyei, Z. (2002). 7. The motivational basis of language learning tasks. In P. Robinson (Ed.), *Language learning & language teaching* (Vol. 2, pp. 137–158). John Benjamins Publishing Company. https://doi.org/10.1075/lllt.2.10dor

Dörnyei, Z. (2005). *The psychology of the language learner: Individual differences in second language acquisition.* (pp. xii, 270). Lawrence Erlbaum Associates Publishers.

Dörnyei, Z. (2007). *Research methods in applied linguistics: Quantitative, qualitative, and mixed methodologies*. Oxford University Press.

Dörnyei, Z. (2019). Towards a better understanding of the L2 learning experience, the Cinderella of the L2 motivational self system. *Studies in Second Language Learning and Teaching*, *9*(1), 19–30. https://doi.org/10.14746/ssllt.2019.9.1.2

Dörnyei, Z., & Kormos, J. (2000). The role of individual and social variables in oral task performance. *Language Teaching Research*, *4*(3), 275–300. https://doi.org/10.1177/136216880000400305

Dörnyei, Z., & Murphey, T. (2008). *Group dynamics in the language classroom* (Nachdr.). Cambridge Univ. Press.

Dörnyei, Z., & Otto, I. (1998). *Motivation in action: A process model of L2 motivation*. https://nottingham-repository.worktribe.com/output/1024190/motivation-in-action-a-process-model-of-l2-motivation

Dörnyei, Z., & Ushioda, E. (Eds.). (2009). *Motivation, language identity and the L2 self*. Multilingual Matters.

Dulay, H. C., & Burt, M. K. (1973). Should we teach children syntax? *Language Learning*, *23*(2), 245–258. https://doi.org/10.1111/j.1467-1770.1973.tb00659.x

Dulay, H. C., & Burt, M. K. (1974). Natural sequences in child second language acquisition. *Language Learning*, *24*(1), 37–53. https://doi.org/10.1111/j.1467-1770.1974.tb00234.x

Eckerth, J. (2008). Investigating consciousness-raising tasks: Pedagogically targeted and non-targeted learning gains. *International Journal of Applied Linguistics*, *18*(2), 119–145. https://doi.org/10.1111/j.1473-4192.2008.00188.x

Ede, L. S., & Lunsford, A. A. (1992). *Singular texts/plural authors: Perspectives on collaborative writing*. Southern Illinois University Press.

Elabdali, R. (2021). Are two heads really better than one? A meta-analysis of the L2 learning benefits of collaborative writing. *Journal of Second Language Writing*, 100788. https://doi.org/10.1016/j.jslw.2020.100788

Ellis, N. (2005). At the interface: Dynamic interactions of explicit and implicit language knowledge. *Studies in Second Language Acquisition, 27*(2), 305-352. https://doi.org/10.1017/S027226310505014X

Ellis, N. (2006). Selective attention and transfer phenomena in L2 acquisition: contingency, cue competition, salience, interference, overshadowing, blocking, and perceptual learning. *Applied Linguistics*, *27*(2), 164–194. https://doi.org/10.1093/applin/aml015

Ellis, N. (2011). Implicit and explicit SLA and their interface. In C. Sanz & R. Leow (Eds.), *Implicit and explicit language learning: conditions, processes, and knowledge in SLA* (pp. 35–47). Georgetown University Press.

Ellis, R. (1994). A theory of instructed second language acquisition. In N. Ellis (Ed.), *Implicit and explicit learning of languages* (pp. 79–114). Academic Press.

Ellis, R. (2003). *Task-based language learning and teaching*. Oxford University Press. http://catalog.hathitrust.org/api/volumes/oclc/52180058.html

Ellis, R. (2009). Task-based language teaching: Sorting out the misunderstandings: Task-based language teaching. *International Journal of Applied Linguistics*, *19*(3), 221–246. https://doi.org/10.1111/j.1473-4192.2009.00231.x

Ellis, R. (2015). *Understanding second language acquisition* (Second edition). Oxford University Press.

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

Ellis, R. (2018). *Reflections on task-based language teaching*. Multilingual Matters.

Ellis, R., & Barkhuizen, G. P. (2005). *Analysing learner language*. Oxford University Press.

Ellis, R., Skehan, P., Li, S., Shintani, N., & Lambert, C. (2020). *Task-based language teaching: Theory and practice*. Cambridge University Press.

Elosua, P. (2005). *Psikometria-testen eraketa eta erabilera*. Euskal Herriko Unibertsitateko Argitalpen Zerbitzua.

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

Etxeberria, F., & Etxeberria, J. (2015). Bilingual education In the Basque Country (1960-2013). In F. Tochon (Ed.), *Language education policy unlimited: global perspectives and local practices* (pp. 249–277). Deep University Press.

Euskal Estatistika Erakundea (EUSTAT) (2021). *Students enrolled in non-university general education in the Basque Country by province and education level, according to centre ownership and language model Advance data. 2020-2021*. https://en.eustat.eus/elementos/ele0002400/ti_alumnado-matriculado-en-ensenanzas-de-regimen-general-no-universitarias-en-la-ca-de-euskadi-por-territorio-historico-y-nivel-de-ensenanza-segun-titularidad-del-centro-y-modelo-linguistico-avance-de-datos-2020-2021/tbl0002427_i.html

European Commission/EACEA/Eurydice. (2017). *Key data on teaching languages at school in Europe – 2017 edition Eurydice report*. Publications Office of the European Union.

Fernández Dobao, A. (2012). Collaborative writing tasks in the L2 classroom: Comparing group, pair, and individual work. *Journal of Second Language Writing*, *21*(1), 40–58. https://doi.org/10.1016/j.jslw.2011.12.002

Fernández Dobao, A., & Blum, A. (2013). Collaborative writing in pairs and small groups: Learners' attitudes and perceptions. *System*, *41*(2), 365–378. https://doi.org/10.1016/j.system.2013.02.002

Fernández, M.-J., Montanero, M., & Lucero, M. (2019). La evaluación de la competencia narrativa en la educación básica. *Revista de Educación*, *383*. https://doi.org/10.4438/1988-592X-RE-2019-383-402

Field, A. (2017). *Discovering statistics using IBM SPSS statistics* (5th edition). SAGE Publications.

Flower, L., & Hayes, J. R. (1981). A cognitive process theory of writing. *College Composition and Communication*, *32*(4), 365. https://doi.org/10.2307/356600

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

Fortune, A., & Thorp, D. (2001). Knotted and entangled: New light on the identification, classification and value of language related episodes in collaborative output tasks. *Language Awareness*, *10*(2–3), 143–160. https://doi.org/10.1080/09658410108667031

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

Foster, P., & Wigglesworth, G. (2016). Capturing accuracy in second language performance: The case for a weighted clause ratio. *Annual Review of Applied Linguistics*, *36*, 98–116. https://doi.org/10.1017/S0267190515000082

Fujii, A., Ziegler, N., & Mackey, A. (2016). 2. Peer interaction and metacognitive instruction in the EFL classroom. In M. Sato & S. Ballinger (Eds.), *Language Learning & Language Teaching* (Vol. 45, pp. 63–89). John Benjamins Publishing Company. https://doi.org/10.1075/lllt.45.03fuj

Gallardo del Puerto, F., & Blanco Suárez, Z. (2021). Foreign language motivation in primary education students: The effects of additional CLIL and gender. *Journal of Immersion and Content-Based Language Education,* *9*(1), 58–84. https://doi.org/10.1075/jicb.19023.gal

Gallego, M. (2014). Second language learners' reflections on the effectiveness of dictogloss: A multi-sectional, multi-level analysis. *Studies in Second Language Learning and Teaching*, *4*(1), 33–50. https://doi.org/10.14746/ssllt.2014.4.1.3

Gallego, M. (2019). L2 Spanish morphosyntactic development through collaborative writing: An analysis of mood production, text length and syntactic complexity. *Language Teaching Research*, 136216881988540. https://doi.org/10.1177/1362168819885406

García Bermejo, M. L. (2021). Foreign language education in Spain: A historical view. *European Journal of Applied Linguistics,* *9*(1), 115–135. https://doi.org/10.1515/eujal-2020-0021

García Hernández, F. J., Roca de Larios, J., & Coyle, Y. (2017). Reformulation as a problem-solving space for young EFL writers: A longitudinal study of language learning strategy. In *Learning foreign languages in primary school: research insights* (pp. 193–222). Multilingual Matters.

García Mayo, M. P. (2002a). Interaction in advanced EFL pedagogy: A comparison of form-focused activities. *International Journal of Educational Research*, *37*(3–4), 323–341. https://doi.org/10.1016/S0883-0355(03)00008-9

García Mayo, M. P. (2002b). The effectiveness of two form-focused tasks in advanced EFL pedagogy. *International Journal of Applied Linguistics*, *12*(2), 156–175. https://doi.org/10.1111/1473-4192.t01-1-00029

García Mayo, M. P. (2012). La formación del profesorado de lenguas extranjeras. In E. Alcón Soler (Ed.), *La educación superior desde una perspectiva multilingüe* (pp. 203-211). Tecnos.

García Mayo, M. P. (Ed.). (2017). *Learning foreign languages in primary school research insights*. Multilingual Matters.

García Mayo, M. P. (2018a). Dictogloss technique. In J. I. Liontas, T. International Association, & M. DelliCarpini (Eds.), *The TESOL encyclopedia of english language teaching* (pp. 1–5). John Wiley & Sons, Inc. https://doi.org/10.1002/9781118784235.eelt0181

García Mayo, M. P. (2018b). Child task-supported interaction in the Spanish EFL setting. Research and challenges. *International Journal of English Studies*, *18*(2), 119–143. https://doi.org/10.6018/ijes/2018/2/319731

García Mayo, M. P. (2021a). Are you coming back? It was fun' Turning ethical and methodological challenges into opportunities in task-based research with children. In A. Pinter & K. Kuchah (Eds.), *Ethical and methodological issues in researching young language learners in school contexts*. Multilingual Matters.

García Mayo, M. P. (Ed.) (2021b). Research on EFL learning by young children in Spain [Special issue]. *Language Teaching for Young Learners, 3*(2).

García Mayo, M. P. (Ed.). (2021c). *Working collaboratively in second/foreign language learning*. De Gruyter. https://doi.org/10.1515/9781501511318

García Mayo, M. P., & Azkarai, A. (2016). EFL task-based interaction: Does task modality impact on language-related episodes? In M. Sato & S. Ballinger (Eds.), *Language learning & language teaching* (Vol. 45, pp. 241–266). John Benjamins Publishing Company. https://doi.org/10.1075/lllt.45.10gar

García Mayo, M. P., & García Lecumberri, M. L. (Eds.). (2003). *Age and the acquisition of English as a foreign language*. Multilingual Matters.

García Mayo, M. P., & Gutiérrez Mangado, M. J. (Eds.) (2021). English language learning in primary schools: Variables at play [Special issue]. *Studies in Second Language Learning and Teaching, 10*(3).

García Mayo, M. P., & Hidalgo, M. Á. (2017). L1 use among young EFL mainstream and CLIL learners in task-supported interaction. *System*, *67*, 132–145. https://doi.org/10.1016/j.system.2017.05.004

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

García Mayo, M. P., & Imaz Agirre, A. (2019). Task modality and pair formation method: Their impact on patterns of interaction and LREs among EFL primary school children. *System*, *80*, 165–175. https://doi.org/10.1016/j.system.2018.11.011

García Mayo, M. P., & Loidi, U. (2017). The use of models as written corrective feedback in English as a foreign language (EFL) Writing. *Annual Review of Applied Linguistics*, *37*, 110–127. https://doi.org/10.1017/S0267190517000071

García Mayo, M. P., & Villarreal Olaizola, I. (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. https://doi.org/10.1177/0267658310386523

García Mayo, M. P., Imaz Agirre, A., & Azkarai, A. (2018). Task repetition effects on CAF in EFL child task-based oral interaction. In M. J. Ahmadian & M. P. García Mayo (Eds.), *Recent perspectives on task-based language learning and teaching* (pp. 11–28). De Gruyter. https://doi.org/10.1515/9781501503399-002

Gardner, R. C., & Lambert, W. E. (1972). *Attitudes and motivation in second-language learning*. Newbury House.

Garrett, P., & Gallego Balsà, L. (2014). International universities and implications of internationalisation for minority languages: Views from university students in Catalonia and Wales. *Journal of Multilingual and Multicultural Development*, *35*(4), 361–375. https://doi.org/10.1080/01434632.2013.874434

Gass, S., & Mackey, A. (2006). Input, Interaction and Output: An Overview. *AILA Review*, *19*, 3–17. https://doi.org/10.1075/aila.19.03gas

Gass, S., Mackey, A., & Ross-Feldman, L. (2005). Task-based Interactions in classroom and laboratory settings: Language Learning Vol. 55, No. 4. *Language Learning*, *55*(4), 575–611. https://doi.org/10.1111/j.0023-8333.2005.00318.x

Genesee, F. (1987). *Learning through two languages: Studies of immersion and bilingual education*. Newbury House Publishers.

Ghanbari, B., Barati, H., & Moinzadeh, A. (2012). Rating scales revisited: EFL Writing assessment context of Iran under scrutiny. *Language Testing in Asia*, *2*(1). https://doi.org/10.1186/2229-0443-2-1-83

Ghorbani, M. (2018). Interactions between type of form-focused instruction, type of morphosyntactic form, and type of language knowledge. *International Review of Applied Linguistics in Language Teaching*, *0*(0). https://doi.org/10.1515/iral-2016-0045

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

Goldschneider, J. M., & DeKeyser, R. (2005). Explaining the “natural order of L2 morpheme acquisition” in English: A meta-analysis of multiple determinants. *Language Learning*, *55*(S1), 27–77. https://doi.org/10.1111/j.0023-8333.2005.00295.x

González-Lloret, M. (2017). Technology and task-based language teaching. In S. L. Thorne & S. May (Eds.), *Language, education and technology* (pp. 193–205). Springer International Publishing. https://doi.org/10.1007/978-3-319-02237-6_16

Gorman, M., & Ellis, R. (2019). The relative effects of metalinguistic explanation and direct written corrective feedback on children’s grammatical accuracy in new writing. *Language Teaching for Young Learners*, *1*(1), 57–81. https://doi.org/10.1075/ltyl.00005.gor

Government of Navarre. (2017). *EPE6 English literacy marking guidelines (2016-2017)*. Sección de Evaluación y calidad. https://www.educacion.navarra.es/documents/27590/1094933/Practice_Test_LRW_Ingles_EP6_PAI_CC.pdf/dd27bd92-4277-f80e-9d77-de0b68733362

Granena, G. (2014). Language aptitude and long-term achievement in early childhood L2 Learners. *Applied Linguistics*, *35*(4), 483–503. https://doi.org/10.1093/applin/amu013

Granena, G., Jackson, D. O., & Yilmaz, Y. (Eds.). (2016). *Cognitive individual differences in second language processing and acquisition*. John Benjamins Publishing Company.

Gris Roca, J. (2017). Grammar in primary school EFL textbook activities: A corpus-driven study of their teaching nature and related cognitive implications. *Research in Corpus Linguistics*, 57–72. https://doi.org/10.32714/ricl.05.05

Guadagnoli, E., & Velicer, W. F. (1988). Relation of sample size to the stability of component patterns. *Psychological Bulletin*, *103*(2), 265–275. https://doi.org/10.1037/0033-2909.103.2.265

Guiraud, P. (1959). *Problèmes et méthodes de la statistique linguistique*. Reidel.

Gunnarsson, C. (2012). Chapter 2.03.03: What Is 'Improvement' in L2 French Writing? In M. Torrance, D. Alamargot, M. Castelló, F. Ganier, O. Kruse, A. Mangen, L. Tolchinsky, & L. van Waes (Eds.), *Learning to write effectively: Current trends in European research* (pp. 225–227). BRILL. https://doi.org/10.1163/9781780529295_051

Halliwell, S. (1992). *Teaching English in the primary classroom*. Longman.

Hanaoka, O., & Izumi, S. (2012). Noticing and uptake: Addressing pre-articulated covert problems in L2 writing. *Journal of Second Language Writing*, *21*(4), 332–347. https://doi.org/10.1016/j.jslw.2012.09.008

Hansen Edwards, J., & Liu, J. (2018). *Peer response in second language writing classrooms, Second Edition*. University of Michigan Press. https://doi.org/10.3998/mpub.9361097

Hansen, J. G. (2005). Guiding principles for effective peer response. *ELT Journal*, *59*(1), 31–38. https://doi.org/10.1093/elt/cci004

Harmer, J. (2013). *The practice of English language teaching: With DVD* (4. ed., 8. impression). Pearson Education.

Henry, A., & Thorsen, C. (2018). Disaffection and agentic engagement: 'Redesigning' activities to enable authentic self-expression. *Language Teaching Research*, 136216881879597. https://doi.org/10.1177/1362168818795976

Hidalgo, M. Á., & García Mayo, M. P. (2019). The influence of task repetition type on young EFL learners' attention to form. *Language Teaching Research*, 136216881986555. https://doi.org/10.1177/1362168819865559

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

Hironymous, P. C. (1993). *Universal principles affecting second language acquisition of the third person singular -s*. George Mason University.

Hirose, K. (2003). Comparing L1 and L2 organizational patterns in the argumentative writing of Japanese EFL students. *Journal of Second Language Writing*, *12*(2), 181–209. https://doi.org/10.1016/S1060-3743(03)00015-8

Horst, M., White, J., & Bell, P. (2010). First and second language knowledge in the language classroom. *International Journal of Bilingualism*, *14*(3), 331–349. https://doi.org/10.1177/1367006910367848

Housen, A., Kuiken, F., & Vedder, I. (2012). Complexity, accuracy and fluency: Definitions, measurement and research. In A. Housen, F. Kuiken, & I. Vedder (Eds.), *Language Learning & Language Teaching* (Vol. 32, pp. 1–20). John Benjamins Publishing Company. https://doi.org/10.1075/lllt.32.01hou

Huang, T., Steinkrauss, R., & Verspoor, M. (2021). Variability as predictor in L2 writing proficiency. *Journal of Second Language Writing*, 100787. https://doi.org/10.1016/j.jslw.2020.100787

Hunt, K. W. (1965). A synopsis of clause-to-sentence length factors. *The English Journal*, *54*(4), 300–309.

IBM Corp. (2016). *IBM SPSS Statistics for Windows, , Version 24.0*. IBM Corp.

Imaz Agirre, A., & García Mayo, M. P. (2013). Gender agreement in L3 English by Basque/Spanish bilinguals. *Linguistic Approaches to Bilingualism*, *3*(4), 415–447. https://doi.org/10.1075/lab.3.4.02ima

Imaz Agirre, A., & García Mayo, M. P. (2018). Chapter 8. Proficiency and transfer effects in the acquisition of gender agreement by L2 and L3 English learners. In J. Cho, M. Iverson, T. Judy, T. Leal, & E. Shimanskaya (Eds.), *Studies in bilingualism* (Vol. 55, pp. 203–227). John Benjamins Publishing Company. https://doi.org/10.1075/sibil.55.08ima

Ishikawa, M. (2018). Written languaging, learners' proficiency levels and L2 grammar learning. *System*, *74*, 50–61. https://doi.org/10.1016/j.system.2018.02.017

Ishikawa, M., & Révész, A. (2020). Chapter 10. L2 learning and the frequency and quality of written languaging. In W. Suzuki & N. Storch (Eds.), *Languaging in Language Learning and Teaching: A collection of empirical studies* (pp. 220–240). John Benjamins Publishing Company. https://doi.org/10.1075/lllt.55.10ish

Jacobs, H. L., Zinkgraf, S., Wormuth, D., Hearfiel, V., & Hughey, J. (1981). *Testing ESL composition: A practical approach. English composition program*. Newbury House Publishers, Inc.

Jarvis, S. (2013). Capturing the diversity in lexical diversity: Lexical diversity. *Language Learning*, *63*, 87–106. https://doi.org/10.1111/j.1467-9922.2012.00739.x

Jarvis, S., & Pavlenko, A. (2008). *Crosslinguistic influence in language and cognition*. Routledge.

JASP Team. (2019). *JASP (Version 0.11.1)[Computer software]*. https://jasp-stats.org/

Jeon, E. Y. (2015). Multiple regression. In L. Plonsky (Ed.), *Advancing quantitative methods in second language acquisition* (pp. 131–158). Routledge.

Johnson, J. S., & Newport, E. L. (1989). Critical period effects in second language learning: The influence of maturational state on the acquisition of English as a second language. *Cognitive Psychology*, *21*(1), 60–99. https://doi.org/10.1016/0010-0285(89)90003-0

Johnson, K. (2008). *An introduction to foreign language learning and teaching*. Pearson Education.

Kail, R., & Salthouse, T. A. (1994). Processing speed as a mental capacity. *Acta Psychologica*, *86*(2–3), 199–225. https://doi.org/10.1016/0001-6918(94)90003-5

Kang, E. Y., Sok, S., & Han, Z. (2019). Thirty-five years of ISLA on form-focused instruction: A meta-analysis. *Language Teaching Research*, *23*(4), 428–453. https://doi.org/10.1177/1362168818776671

Kang, S., & Lee, J.-H. (2019). Are two heads always better than one? The effects of collaborative planning on L2 writing in relation to task complexity. *Journal of Second Language Writing*, *45*, 61–72. https://doi.org/10.1016/j.jslw.2019.08.001

Kasprowicz, R., & Marsden, E. (2018). Towards ecological validity in research into input-based practice: Form spotting can be as beneficial as form-meaning practice. *Applied Linguistics*, *39*(6), 886–911. https://doi.org/10.1093/applin/amw051

Khezrlou, S. (2019). Effects of timing and availability of isolated FFI on learners' written accuracy and fluency through task repetition. *The Language Learning Journal*, 1–13. https://doi.org/10.1080/09571736.2019.1656765

Khezrlou, S. (2020). The role of task repetition with direct written corrective feedback in L2 writing complexity, accuracy and fluency. *Journal of Second Language Studies*, *3*(1), 31–54. https://doi.org/10.1075/jsls.19025.khe

Khezrlou, S. (2021). Explicit instruction through task repetition: Effects on explicit and implicit knowledge development. *Language Awareness*, 1–23. https://doi.org/10.1080/09658416.2020.1866590

Kihlstedt, M. (2019). Foreign Language (FL) Teaching and Learning in Primary Schools in Europe: Beliefs and Realities. *Emerging Trends in Education*, *2*(3). https://doi.org/10.19136/etie.a2n3.3243

Kim, Y., Choi, B., Yun, H., Kim, B., & Choi, S. (2020). Task repetition, synchronous written corrective feedback and the learning of Korean grammar: A classroom-based study. *Language Teaching Research*, 136216882091235. https://doi.org/10.1177/1362168820912354

Kim, Y., Kang, S., Yun, H., Kim, B., & Choi, B. (2020). The role of task repetition in a Korean as a foreign language classroom: Writing quality, attention to form, and learning of Korean grammar. *Foreign Language Annals*, *53*(4), 827–849. https://doi.org/10.1111/flan.12501

Kim, Y., & McDonough, K. (2008). The effect of interlocutor proficiency on the collaborative dialogue between Korean as a second language learners. *Language Teaching Research*, *12*(2), 211–234. https://doi.org/10.1177/1362168807086288

Kim, Y., & McDonough, K. (2011). Using pretask modelling to encourage collaborative learning opportunities. *Language Teaching Research*, *15*(2), 183–199. https://doi.org/10.1177/1362168810388711

Kline, R. B. (2011). *Principles and practice of structural equation modeling, 3rd ed.* (pp. xvi, 427). Guilford Press.

Kopinska, M., & Azkarai, A. (2020). Exploring young EFL learners' motivation: Individual versus pair work on dictogloss tasks. *Studies in Second Language Learning and Teaching*, *10*(3), 607–630. https://doi.org/10.14746/ssllt.2020.10.3.10

Kormos, J. (2011). Task complexity and linguistic and discourse features of narrative writing performance. *Journal of Second Language Writing*, *20*(2), 148–161. https://doi.org/10.1016/j.jslw.2011.02.001

Kormos, J., & Dörnyei, Z. (2004). The interaction of linguistic and motivational variables in second language task performance. *Zeitschrift Für Interkulturellen Fremdsprachenunterricht*, *9*(2), Online.

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

Kowal, M., & Swain, M. (1997). From semantic to syntactic processing: How can we promote it in the immersion classroom? In R. K. Johnson & M. Swain (Eds.), *Immersion Education* (pp. 284–310). Cambridge University Press. https://doi.org/10.1017/CBO9781139524667.022

Krashen, S. D. (1982). Acquiring a second language. *World Englishes*, *1*(3), 97–101. https://doi.org/10.1111/j.1467-971X.1982.tb00476.x

Krashen, S. D. (1983). Newmark's ignorance hypothesis and current second language acquisition theory. In S. Gass & L. Selinker (Eds.), *Language transfer in language learning* (pp. 135–153). Newbury House.

Krashen, S. D., Long, M., & Scarcella, R. C. (1979). Age, rate and eventual attainment in second language acquisition. *TESOL Quarterly*, *13*(4), 573. https://doi.org/10.2307/3586451

Kuhl, P. K. (2010). Brain mechanisms in early language acquisition. *Neuron*, *67*(5), 713–727. https://doi.org/10.1016/j.neuron.2010.08.038

Kuiken, F., & Vedder, I. (2002a). Collaborative writing in L2: The effect of group interaction on text quality. In S. Ransdell & M.-L. Barbier (Eds.), *New directions for research in L2 writing* (Vol. 11, pp. 169–188). Springer Netherlands. https://doi.org/10.1007/978-94-010-0363-6_9

Kuiken, F., & Vedder, I. (2002b). The effect of interaction in acquiring the grammar of a second language. *International Journal of Educational Research*, *37*(3–4), 343–358. https://doi.org/10.1016/S0883-0355(03)00009-0

Kuiken, F., & Vedder, I. (2014). Rating written performance: What do raters do and why? *Language Testing*, *31*(3), 329–348. https://doi.org/10.1177/0265532214526174

Kuiken, F., & Vedder, I. (2018). Chapter 11. Assessing functional adequacy of L2 performance in a task-based approach. In N. Taguchi & Y. Kim (Eds.), *Task-based language teaching* (Vol. 10, pp. 266–285). John Benjamins Publishing Company. https://doi.org/10.1075/tblt.10.11kui

Kyle, K., & Crossley, S. A. (2018). Measuring syntactic complexity in L2 writing using fine-grained clausal and phrasal indices. *The Modern Language Journal*, *102*(2), 333–349. https://doi.org/10.1111/modl.12468

Lahuerta Martínez, A. C. (2017). Syntactic complexity in secondary-level English writing: differences among writers enrolled on bilingual and non-bilingual programmes. *Porta Linguarum Revista Interuniversitaria de Didáctica de Las Lenguas Extranjeras*. https://doi.org/10.30827/Digibug.54003

Lamb, M. (2004). Integrative motivation in a globalizing world. *System, 32*(1), 3–19. https://doi.org/10.1016/j.system.2003.04.002

Lambert, C. (2017). Tasks, affect and second language performance. *Language Teaching Research*, *21*(6), 657–664. https://doi.org/10.1177/1362168817736644

Lantolf, J. (Ed.). (2000). *Sociocultural theory and second language learning*. Oxford Univ. Press.

Lantolf, J. (2012). Sociocultural theory: A dialectical approach to L2 research. In S. Gass & A. Mackey (Eds.), *Handbook of second language acquisition* (pp. 57–72). Routledge.

Lantolf, J., & Thorne, S. L. (2006). *Sociocultural theory and the genesis of second language development*. Oxford University Press.

LaPierre, D. (1994). *Language output in a cooperative learning setting: Determining its effects on second language learning* [Unpublished MA Thesis]. University of Toronto.

Larsen-Freeman, D. (2006). The emergence of complexity, fluency, and accuracy in the oral and written production of five Chinese learners of English. *Applied Linguistics*, *27*(4), 590–619. https://doi.org/10.1093/applin/aml029

Larson-Hall, J. (2016). *A guide to doing statistics in second language research using SPSS and R* (Second Edition). Routledge.

Larson-Hall, J., & Plonsky, L. (2015). Reporting and interpreting quantitative research findings: What gets reported and recommendations for the field. *Language Learning*, *65*(S1), 127–159. https://doi.org/10.1111/lang.12115

Lasagabaster, D. (2008). Foreign language competence in content and language integrated courses. *The Open Applied Linguistics Journal*, *1*(1), 30–41. https://doi.org/10.2174/1874913500801010030

Lasagabaster, D., & Doiz, A. (2003). Maturational constraints on foreign language written production. In M. P. García Mayo & M. L. García Lecumberri (Eds.), *Age and the acquisition of English as a foreign language* (pp. 136–160). Multilingual Matters.

Lázaro Ibarrola, A., & Azpilicueta-Martinez, R. (2019). Negotiation of meaning in child-child vs. adult-adult interactions: Evidence from low proficiency EFL learners. *International Review of Applied Linguistics in Language Teaching*, *0*(0). https://doi.org/10.1515/iral-2019-0013

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

Lee, I. (2016). 5. EFL writing in schools. In R. M. Manchón & P. K. Matsuda (Eds.), *Handbook of Second and Foreign Language Writing*. De Gruyter. https://doi.org/10.1515/9781614511335-008

Lee, J. (2020). Effects of linguistic and affective variables on middle school students' writing performance in the context of English as a foreign language: An approach using structural equation modeling. *Reading and Writing*, *33*(5), 1235–1262. https://doi.org/10.1007/s11145-019-10007-2

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

Lei, L., & Liu, D. (2019). Research trends in applied linguistics from 2005 to 2016: A bibliometric analysis and its implications. *Applied Linguistics*, *40*(3), 540–561. https://doi.org/10.1093/applin/amy003

Lenneberg, E. H. (1967). *Biological foundations of language*. Wiley.

Leonet, O., Cenoz, J., & Gorter, D. (2019). Developing morphological awareness across languages: Translanguaging pedagogies in third language acquisition. *Language Awareness*, 1–19. https://doi.org/10.1080/09658416.2019.1688338

Leontiev, A. (1981). *Psychology and the language learning process.* Pergamon.

Li, Y., Han, Y., & Gao, X. (2018). Young learners' motivation for learning English. In S. Garton & F. Copland (Eds.), *The Routledge handbook of teaching English to young learners* (pp. 60–72). Routledge.

Lichtman, K. (2016). Age and learning environment: Are children implicit second language learners? *Journal of Child Language*, *43*(3), 707–730. https://doi.org/10.1017/S0305000915000598

Lin, O. P., & Maarof, N. (2013). Collaborative writing in summary writing: student perceptions and problems. *Procedia - Social and Behavioral Sciences*, *90*, 599–606. https://doi.org/10.1016/j.sbspro.2013.07.131

Loewen, S., Erlam, R., & Ellis, R. (2009). *The incidental acquisition of third Person -s as implicit and explicit knowledge*. In R. Ellis, S. Loewen, C. Elder, H. Reinders, R. Erlam, & J. Philp (Eds.) *Implicit and explicit knowledge in second language learning, testing and teaching* (pp. 262–280). Multilingual Matters. https://www.degruyter.com/view/book/9781847691767/10.21832/9781847691767-013.xml

Loewen, S., & Gönülal, T. (2015). Exploratory factor analysis and principal components analysis. In L. Plonsky (Ed.), *Advancing quantitative methods in second language acquisition* (pp. 182–212). Routledge.

Loewen, S., & Plonsky, L. (2015). *An A [to] Z of applied linguistics research methods*. Palgrave Macmillan.

Long, M. (1983). Native speaker/non-native speaker conversation and the negotiation of comprehensible input1. *Applied Linguistics*, *4*(2), 126–141. https://doi.org/10.1093/applin/4.2.126

Long, M. (1988). Instructed interlanguage development. In L. Beebe (Ed.), *Issues in second language acquisition: Multiple perspectives* (pp. 115–141). Newbury House.

Long, M. (1991). Focus on form: A design feature in language teaching methodology. In K. de Bot, R. B. Ginsberg, & C. Kramsch (Eds.), *Studies in bilingualism* (Vol. 2, p. 39). John Benjamins Publishing Company. https://doi.org/10.1075/sibil.2.07lon

Long, M. (1996). The role of the linguistic environment in second language acquisition. In W. Ritchie & T. Bhatia (Eds.), *Handbook of second language acquisition* (pp. 121–158). Academic Press.

Long, M. (2007). *Problems in SLA*. L. Erlbaum Associates.

Long, M. (2015). *Second language acquisition and task-based language teaching* (First Edition). Wiley-Blackwell.

Long, M. (2016). In defense of tasks and TBLT: Nonissues and real issues. *Annual Review of Applied Linguistics*, *36*, 5–33. https://doi.org/10.1017/S0267190515000057

Loschky, L., & Bley-Vroman, R. (1993). Grammar and task-based methodology. *Tasks and Language Learning*, *1*, 123–167.

Lu, X. (2011). A corpus-based evaluation of syntactic complexity measures as indices of college-level ESL writers' language development. *TESOL Quarterly*, *45*(1), 36–62. https://doi.org/10.5054/tq.2011.240859

Luk, Z. P., & Shirai, Y. (2009). Is the acquisition order of grammatical morphemes impervious to L1 knowledge? Evidence from the acquisition of plural - *s* , articles, and possessive ' *s*. *Language Learning*, *59*(4), 721–754. https://doi.org/10.1111/j.1467-9922.2009.00524.x

Luquin, M., & García Mayo, M. P. (2021). Exploring the use of models as a written corrective feedback technique among EFL children. *System*, *98*, 102465. https://doi.org/10.1016/j.system.2021.102465

Lyster, R. (2001). Negotiation of form, recasts, and explicit correction in relation to error types and learner repair in immersion classrooms. *Language Learning*, *51*, 265–301. https://doi.org/10.1111/j.1467-1770.2001.tb00019.x

Lyster, R. (2007). *Learning and teaching languages through content: A counterbalanced approach*. John Benjamins Pub.

Lyster, R., & Ranta, L. (1997). Corrective feedback and learner uptake: Negotiation of form in communicative classrooms. *Studies in Second Language Acquisition*, *19*(1), 37–66. https://doi.org/10.1017/S0272263197001034

Macaro, E. (2005). Codeswitching in the L2 classroom: A communication and learning strategy. In E. Llurda (Ed.), *Non-native language teachers: Perceptions, challenges and contributions to the profession* (pp. 63–84). Springer US. https://doi.org/10.1007/0-387-24565-0_5

Mackey, A., Oliver, R., & Leeman, J. (2003). Interactional input and the incorporation of feedback: An exploration of NS-NNS and NNS-NNS Adult and child dyads. *Language Learning*, *53*(1), 35–66. https://doi.org/10.1111/1467-9922.00210

MacWhinney, B. (2000). *The CHILDES project: Tools for analyzing talk* (3rd edition) [Computer software]. Lawrence Erlbaum.

Malmqvist, A. (2005). How does group discussion in reconstruction tasks affect written language output? *Language Awareness*, *14*(2–3), 128–141. https://doi.org/10.1080/09658410508668829

Malvern, D., Richards, B., Chipere, N., & Durán, P. (2004). *Lexical diversity and language development*. Palgrave Macmillan UK. https://doi.org/10.1057/9780230511804

Manchón, R. M. (2011). Writing to learn the language: Issues in theory and research. In R. M. Manchón (Ed.), *Learning-to-write and writing-to-learn in an additional language* (pp. 61–82). John Benjamins Publishing Company. https://doi.org/10.1075/lllt.31.07man

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

Manchón, R. M. (2014b). The internal dimension of tasks: The interaction between task factors and learner factors in bringing about learning through writing. In H. Byrnes & R. M. Manchón (Eds.), *Task-based language teaching - Insights from L2 writing* (pp. 27–52). John Benjamins Publishing Company. https://doi.org/10.1075/tblt.7.02man

Manchón, R. M., Larios, J. R. D., & Murphy, L. (2009). Chapter 4. The temporal dimension and problem-solving nature of foreign language composing processes. implications for theory. In R. M. Manchón (Ed.), *Writing in foreign language contexts* (pp. 102–129). Multilingual Matters. https://doi.org/10.21832/9781847691859-008

Manchón, R. M., & Vasylets, O. (2019). Language learning through writing: theoretical perspectives and empirical evidence. In J. W. Schwieter & A. Benati (Eds.), *The Cambridge handbook of language learning* (1st ed., pp. 341–362). Cambridge University Press. https://doi.org/10.1017/9781108333603.015

Manchón, R. M., & Williams, J. (2016). 26. L2 writing and SLA studies. In R. M. Manchón & P. K. Matsuda (Eds.), *Handbook of second and foreign language writing*. De Gruyter. https://doi.org/10.1515/9781614511335-029

Marsden, E., Mackey, A., & Plonsky, L. (2016). The IRIS repository: Advancing research practice and methodology. In A. Mackey & E. Marsden (Eds.), *Advancing methodology and practice: The IRIS repository of instruments for research into second languages* (pp. 1–21). Routledge.

Martínez Adrián, M., & Arratibel Irazusta, I. (2020). The interface between task-modality and the use of previously known languages in young CLIL English learners. *Studies in Second Language Learning and Teaching*, *10*(3), 473–500. https://doi.org/10.14746/ssllt.2020.10.3.4

Martínez Adrián, M., & Gutiérrez Mangado, M. J. (2015). Is CLIL instruction beneficial in terms of general proficiency and specific areas of grammar? *Journal of Immersion and Content-Based Language Education*, *3*(1), 51–76. https://doi.org/10.1075/jicb.3.1.03adr

Mateo Cutillas, N. (2016). Teaching writing in the primary school. *Encuentro*, *25*, 45–61.

Matsuzaki-Carreira, J. (2006). Motivation for learning English as a foreign language in Japanese elementary schools. *JALT Journal, 28*(2), 135–158.

McDonough, K., & García Fuentes, C. (2015). The effect of writing task and task conditions on Colombian EFL learners' language use. *TESL Canada Journal*, *32*(2), 67. https://doi.org/10.18806/tesl.v32i2.1208

McDonough, K., & Hernández González, T. (2019). Collaborative prewriting discussions and ESL students' vocabulary development: LREs and depth of engagement. In M. Sato & S. Loewen (Eds.), *Evidence-based second language pedagogy* (pp. 105–122). Routledge.

McDonough, K., De Vleeschauwer, J., & Crawford, W. J. (2018). Exploring the benefits of collaborative prewriting in a Thai EFL context. *Language Teaching Research*, 136216881877352. https://doi.org/10.1177/1362168818773525

Meara, P. M., & Miralpeix, I. (2017). *Tools for researching vocabulary*. Multilingual Matters.

Mercer, N. (1995). *The guided construction of knowledge: Talk amongst teachers and learners.* (pp. viii, 135). Multilingual Matters.

Mercer, S., Ryan, S., & Williams, M. (Eds.). (2012). *Psychology for language learning: Insights from research, theory and practice*. Palgrave Macmillan.

Michel, M. (2017). *Complexity, accuracy, and fluency in L2 production*. In S. Loewen & M. Sato (Eds.), *The Routledge handbook of instructed second language acquisition* (pp. 50–68). Routledge. https://doi.org/10.4324/9781315676968-4

Michel, M., Révész, A., Lu, X., Kourtali, N.-E., Lee, M., & Borges, L. (2020). Investigating L2 writing processes across independent and integrated tasks: A mixed-methods study. *Second Language Research*, *36*(3), 307–334. https://doi.org/10.1177/0267658320915501

Mihaljević Djigunović, J., & Nikolov, M. (2019). Motivation of young learners of foreign languages. In M. Lamb, K. Csizér, A. Henry, & S. Ryan (Eds.), *The Palgrave handbook of motivation for language learning* (pp. 515–533). Springer International Publishing. https://doi.org/10.1007/978-3-030-28380-3_25

Ministerio de Educación, Cultura y Deporte. (2002). *Evaluación de la enseñanza y el aprendizaje de la lengua inglesa: Educación primaria 1999*. Ministerio de Educación Cultura y Deporte de España. https://www.educacionyfp.gob.es/dctm/ievaluacion/nacional/06evaluacion-de-la-ensenanza-y-el-aprendizaje-de-la-lengua-inglesa.-educacion-primaria-1999---avance-de-resultados.pdf?documentId=0901e72b80110dbf

Ministerio de Educación, Cultura y Deporte. (2004). *El objetivo de este estudio es conocer y valorar las capacidades básicas obtenidas por los alumnos de 2º y 4º de la educación secundaria obligatoria en el área de la lengua inglesa*. Ministerio de Educación Cultura y Deporte de España. https://sede.educacion.gob.es/publiventa/d/11674/19/0

Mostafa, T., & Kim, Y. (2020). The effects of input and output based instruction on the development of L2 explicit and automatised explicit knowledge: A classroom based study. *Language Awareness*, 1–25. https://doi.org/10.1080/09658416.2020.1760292

Mozaffari, S. H. (2017). Comparing student-selected and teacher-assigned pairs on collaborative writing. *Language Teaching Research*, *21*(4), 496–516. https://doi.org/10.1177/1362168816641703

Muñoz, C. (1994). A case of frequency-based markedness. *Atlantis*, *XVI*(1–2), 165–177.

Muñoz, C. (Ed.). (2006). *Age and the rate of foreign language learning*. Multilingual Matters.

Muñoz, C. (2007). CLIL: Some thoughts on its psycholinguistic principles. *Revista Española de Lingüística Aplicada (RESLA),* monographic volume, 17-26.

Muñoz, C. (2008). Symmetries and asymmetries of age effects in naturalistic and instructed L2 learning. *Applied Linguistics*, *29*(4), 578–596. https://doi.org/10.1093/applin/amm056

Murphy, V. A. (2014). *Second language learning in the early school years: Trends and contexts*. Oxford University Press.

Mylläri, T. (2020). Words, clauses, sentences, and T-units in learner language: Precise and objective units of measure? *Journal of the European Second Language Association*, *4*(1), 13–23. https://doi.org/10.22599/jesla.63

Nadeau, M., Arseneau, R., Fisher, C., & Lacasse, C. Q. (2020). Les dictées métacognitives interactives: In J. Thibeault & C. Fleuret (Eds.), *Didactique du français en contextes minoritaires* (pp. 155–184). University of Ottawa Press; JSTOR. https://doi.org/10.2307/j.ctv1b9f54d.12

Nadeau, M., Quevillon Lacasse, C., Giguère, M.-H., Arseneau, R., & Fisher, C. (2020). Teaching syntax and punctuation in French L1: How the notion of sentence was operationalized in innovative didactic devices. *L1 Educational Studies in Language and Literature*, *20, Running Issue*, 1–28. https://doi.org/10.17239/L1ESLL-2020.20.03.05

Navés, T., Miralpeix, I., & Luz Celaya, M. (2005). Who transfers more… And what? Crosslinguistic influence in relation to school grade and language dominance in EFL. *International Journal of Multilingualism*, *2*(2), 113–134. https://doi.org/10.1080/14790710508668380

Newmark, L. (1966). How not to interfere in language learning. *International Journal of American Linguistics*, *32*(1), 77–87.

Newton, J., & Le Diem Bui, T. (2020). *3 Low-proficiency learners and task-based language teaching*. In C. Lambert & R. Oliver (Eds.), *Using tasks in second language teaching* (pp. 33–48). Multilingual Matters. https://doi.org/10.21832/9781788929455-004

Nguyen, B. T. T., & Newton, J. (2019). Learner proficiency and EFL learning through task rehearsal and performance. *Language Teaching Research*, https://doi.org/10.1177/1362168818819021

Nitta, R., & Baba, K. (2014). Task repetition and L2 writing development: A longitudinal study from a Dynamic Systems perspective. In H. Byrnes & R. M. Manchón (Eds.), *Task-based language teaching - Insights from and for L2 writing*. John Benjamins Publishing Company. https://doi.org/10.1075/tblt.7.05nit

Niu, R., Jiang, L., & Deng, Y. (2018). Effect of proficiency pairing on l2 learners' language learning and scaffolding in collaborative writing. *The Asia-Pacific Education Researcher*, *27*(3), 187–195. https://doi.org/10.1007/s40299-018-0377-2

Norris, J. M., & Ortega, L. (2000). Effectiveness of L2 instruction: A research synthesis and quantitative meta-analysis. *Language Learning*, *50*(3), 417–528. https://doi.org/10.1111/0023-8333.00136

Norris, J. M., & Ortega, L. (2009). Towards an organic approach to investigating CAF in Instructed SLA: The case of complexity. *Applied Linguistics*, *30*(4), 555–578. https://doi.org/10.1093/applin/amp044

Núñez Cortés, J. A., & Barrado Mendo, A. (2020). Modelos de enseñanza de la escritura en los libros de texto de educación primaria. *Aula de Encuentro*, *22*(2), 6–26. https://doi.org/10.17561/ae.v22n2.5578

Ohta, A. (2001). *Second Language Acquisition Processes in the Classroom: Learning Japanese*. Lawrence Erlbaum. http://www.journals.cambridge.org/abstract_S0272263103220180

Oliver, R. (1998). Negotiation of meaning in child interactions. *The Modern Language Journal*, *82*(3), 372–386. https://doi.org/10.1111/j.1540-4781.1998.tb01215.x

Oliver, R. (2002). The patterns of negotiation for meaning in child interactions. *The Modern Language Journal*, *86*(1), 97–111. https://doi.org/10.1111/1540-4781.00138

Oliver, R. (2009). How young is too young? Investigating negotiation of meaning and corrective feedback in children aged five to seven years. In A. Mackey & C. Polio (Eds.), *Multiple perspectives on interaction: second language research in honor of Susan M. Gass* (pp. 135–156).

Oliver, R., Nguyen, B., & Sato, M. (2017). Child SLA. In S. Loewen & M. Sato (Eds.), *The Routledge handbook of instructed second language acquisition* (pp. 468–487). Routledge.

Ong, J., & Zhang, L. J. (2010). Effects of task complexity on the fluency and lexical complexity in EFL students' argumentative writing. *Journal of Second Language Writing*, *19*(4), 218–233. https://doi.org/10.1016/j.jslw.2010.10.003

Ong, J., & Zhang, L. J. (2013). Effects of the manipulation of cognitive processes on EFL writers' text quality. *TESOL Quarterly*, *47*(2), 375–398. https://doi.org/10.1002/tesq.55

Order of 27th April 2016. (2016). *Boletín Oficial del País Vasco,* N 79-1765 https://www.euskadi.eus/bopv2/datos/2016/04/1601765a.pdf

Organic Law 2/2006 of 3rd May, of Education (LOE) (2006). *Boletín Oficial del Estado,* 4th May 2006.

Organic Law 8/2013 of 9th December for the education quality improvement (LOMCE) (2013). *Boletín Oficial del Estado,* 10th December 2013. https://www.boe.es/buscar/pdf/2013/BOE-A-2013-12886-consolidado.pdf

Organic Law 3/2020, of 29th December for modifying the basic regulating framework of the current non-university education law in Spain (LOMLOE) (2020). *Boletín Oficial del Estado,* 30th December 2020. https://www.boe.es/eli/es/lo/2020/12/29/3/dof/spa/pdf

Ortega, L. (2009). Studying writing across EFL contexts: Looking back and moving forward. In R. M. Manchón (Ed.), *Writing in foreign language contexts: learning, teaching and research* (pp. 232–255). Multilingual Matters.

Ortega, Lourdes. (2012). Epilogue: Exploring L2 writing–SLA interfaces. *Journal of Second Language Writing*, *21*(4), 404–415. https://doi.org/10.1016/j.jslw.2012.09.002

Pae, T.-I. (2008). A structural model for Korean EFL students' writing performance. *English Teaching*, *63*(2), 121–137.

Pae, T.-I. (2018). Effects of task type and L2 proficiency on the relationship between L1 and L2 in reading and writing. *Studies in Second Language Acquisition*, *40*(01), 63–90. https://doi.org/10.1017/S0272263116000462

Pallotti, G. (2009). CAF: Defining, refining and differentiating constructs. *Applied Linguistics*, *30*(4), 590–601. https://doi.org/10.1093/applin/amp045

Pawlak, M. (Ed.). (2016). *Classroom-oriented research: reconciling theory and practice*. Springer International Publishing. https://doi.org/10.1007/978-3-319-30373-4

Payant, C. (2020). Exploring multilingual learners' writing practices during an L2 and an L3 individual writing task. *The Canadian Modern Language Review*, *76*(4), 313–334. https://doi.org/10.3138/cmlr-2020-0030

Payant, C., & Kim, Y. (2019). Impact of task modality on collaborative dialogue among plurilingual learners: A classroom-based study. *International Journal of Bilingual Education and Bilingualism*, *22*(5), 614–627. https://doi.org/10.1080/13670050.2017.1292999

Peng, J., Wang, C., & Lu, X. (2020). Effect of the linguistic complexity of the input text on alignment, writing fluency, and writing accuracy in the continuation task. *Language Teaching Research*, *24*(3), 364–381. https://doi.org/10.1177/1362168818783341

Philp, J., & Duchesne, S. (2016). Exploring engagement in tasks in the language classroom. *Annual Review of Applied Linguistics*, *36*, 50–72. https://doi.org/10.1017/S0267190515000094

Philp, J., Adams, R. J., & Iwashita, N. (2014). *Peer interaction and second language learning*. Routledge, Taylor & Francis Group.

Philp, J., Mackey, A., & Oliver, R. (2008). Child's play? Second language acquisition and the younger learner in context. In J. Philp, R. Oliver, & A. Mackey (Eds.), *Second language acquisition and the younger learner* (pp. 3–23). John Benjamins Publishing Company. https://doi.org/10.1075/lllt.23.02phi

Piaget, J. (1926/1955). *The language and thought of the child*. Routledge and Kegan Paul.

Pica, T. (1983). Methods of morpheme quantification: Their effect on the interpretation of second language data. *Studies in Second Language Acquisition*, *6*(1), 69–78.

Pica, T., & Doughty, C. (1985). Input and interaction in the communicative language classroom: A comparison of teacher-fronted and group activities. In S. Gass & C. G. Madden (Eds.), *Input in second language acquisition* (pp. 115–132). Newbury House.

Pienemann, M. (1998). *Language processing and second language development: processability theory* (Vol. 15). John Benjamins Publishing Company. https://doi.org/10.1075/sibil.15

Pienemann, M. (Ed.). (2005). *Cross-linguistic aspects of processability theory* (Vol. 30). John Benjamins Publishing Company. https://doi.org/10.1075/sibil.30

Pinter, A. (2006). Verbal evidence of task related strategies: Child versus adult interactions. *System*, *34*(4), 615–630. https://doi.org/10.1016/j.system.2006.09.005

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

Pinter, A. (2007b). What children say: Benefits of task repetition. In K. Van den Branden, K. Van Gorp, & M. Verhelst (Eds.), *Tasks in action: task-based language education from a classroom-based perspective* (pp. 131–158). Cambridge Scholars Publishing.

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

Pinter, A. (2015). Task-based learning with children. In J. Bland (Ed.), *Teaching English to young learners: Critical issues in language teaching with 3-12 year olds* (pp. 113–128). Bloomsbury Academic, an imprint of Bloomsbury Publishing. http://www.worldcat.org/oclc/897400059

Pinter, A. (2018). Research issues with young learners. In S. Garton & F. Copland (Eds.), *The Routledge handbook of teaching English to young learners* (pp. 411–424). Routledge.

Pinter, A., & Kuchah, K. (Eds.). (2021). *Ethical and methodological issues in researching young language learners in school contexts*. Multilingual Matters.

Pladevall-Ballester, E. (2019). A longitudinal study of primary school EFL learning motivation in CLIL and non-CLIL settings. *Language Teaching Research, 23*(6), 765–786. https://doi.org/10.1177/1362168818765877

Pladevall-Ballester, E., & Vraciu, A. (2020). EFL child peer interaction: Measuring the effect of time, proficiency pairing and language of interaction. *Studies in Second Language Learning and Teaching*, *10*(3), 449–472. https://doi.org/10.14746/ssllt.2020.10.3.3

Plonsky, L. (2013). Study quality in SLA: An assessment of designs, analyses, and reporting practices in quantitative L2 research. *Studies in Second Language Acquisition*, *35*(4), 655–687. https://doi.org/10.1017/S0272263113000399

Plonsky, L., & Derrick, D. J. (2016). A meta-analysis of reliability coefficients in second language research. *The Modern Language Journal*, *100*(2), 538–553. https://doi.org/10.1111/modl.12335

Plonsky, L., & Ghanbar, H. (2018). Multiple regression in L2 research: A methodological synthesis and guide to interpreting R^2 values. *The Modern Language Journal*, *102*(4), 713–731. https://doi.org/10.1111/modl.12509

Plonsky, L., & Gönülal, T. (2015). Methodological synthesis in quantitative L2 research: A review of reviews and a case study of exploratory factor analysis: Methodological reviews and a case study of EFA. *Language Learning*, *65*(S1), 9–36. https://doi.org/10.1111/lang.12111

Plonsky, L., & Kim, Y. (2016). Task-based learner production: A substantive and methodological review. *Annual Review of Applied Linguistics*, *36*, 73–97. https://doi.org/10.1017/S0267190516000015

Plonsky, L., & Oswald, F. L. (2014). How big is “big”? Interpreting effect sizes in L2 research: effect sizes in L2 research. *Language Learning*, *64*(4), 878–912. https://doi.org/10.1111/lang.12079

Polio, C. (2012). The relevance of second language acquisition theory to the written error correction debate. *Journal of Second Language Writing*, *21*(4), 375–389. https://doi.org/10.1016/j.jslw.2012.09.004

Polio, C. (2019). Keeping the language in second language writing classes. *Journal of Second Language Writing*, *46*, 100675. https://doi.org/10.1016/j.jslw.2019.100675

Polio, C., & Friedman, D. (2016). *Understanding, evaluating, and conducting second language writing research* (1st ed.). Routledge. https://doi.org/10.4324/9781315747293

Polio, C., & Shea, M. C. (2014). An investigation into current measures of linguistic accuracy in second language writing research. *Journal of Second Language Writing*, *26*, 10–27. https://doi.org/10.1016/j.jslw.2014.09.003

Porral Combarros, L. (2017). *Teaching EFL through task repetition and collaborative writing* [Master's Thesis]. http://hdl.handle.net/11162/174475

Qiu, X., & Lee, M.-K. (2020). Regulated learning and self-efficacy beliefs in peer collaborative writing: An exploratory study of L2 learners' written products, task discussions, and self-reports. *System*, *93*, 102312. https://doi.org/10.1016/j.system.2020.102312

Qiu, X., & Lo, Y. Y. (2017). Content familiarity, task repetition and Chinese EFL learners' engagement in second language use. *Language Teaching Research*, *21*(6), 681–698. https://doi.org/10.1177/1362168816684368

QSR International. (2012). *NVivo 10* (Version 10) [Computer software].

Ranta, L. (2002). 8. The role of learners' language analytic ability in the communicative classroom. In P. Robinson (Ed.), *Language Learning & Language Teaching* (Vol. 2, pp. 159–180). John Benjamins Publishing Company. https://doi.org/10.1075/lllt.2.11ran

Read, C., & Ormerod, M. (2018). *New Tiger 6*. Macmillan Education UK.

Révész, A. (2019). Replication in task-based language teaching research: Kim (2012) and Shintani (2012). *Language Teaching*, *52*(3), 374–384. https://doi.org/10.1017/S0261444817000283

Révész, A., Kourtali, N.-E., & Mazgutova, D. (2017). Effects of task complexity on L2 writing behaviors and linguistic complexity: Task complexity and L2 writing. *Language Learning*, *67*(1), 208–241. https://doi.org/10.1111/lang.12205

Riazi, M., Shi, L., & Haggerty, J. (2018). Analysis of the empirical research in the journal of second language writing at its 25th year (1992–2016). *Journal of Second Language Writing*, *41*, 41–54. https://doi.org/10.1016/j.jslw.2018.07.002

Ringbom, H. (2007). *Cross-linguistic similarity in foreign language learning*. Multilingual Matters.

Rinnert, C., & Kobayashi, H. (2016). 17. Multicompetence and multilingual writing. In R. M. Manchón & P. K. Matsuda (Eds.), *Handbook of second and foreign language writing*. De Gruyter. https://doi.org/10.1515/9781614511335-020

Rixon, S. (2018). Teaching English to young learners in Europe. In S. Garton & F. Copland (Eds.), *The Routledge handbook of teaching English to young learners* (pp. 493–507). Routledge.

Robinson, P. (2001). Task complexity, cognitive resources, and syllabus design: A triadic framework for examining task influences on SLA. In P. Robinson (Ed.), *Cognition and Second Language Instruction* (1st ed., pp. 287–318). Cambridge University Press. https://doi.org/10.1017/CBO9781139524780.012

Roehr-Brackin, K., & Tellier, A. (2019). The role of language-analytic ability in children's instructed second language learning. *Studies in Second Language Acquisition*, *41*(5), 1111–1131. https://doi.org/10.1017/S0272263119000214

Roos, J. (2019). Chapter 12. Exploiting the potential of tasks for targeted language learning in the EFL classroom. In A. Lenzing, H. Nicholas, & J. Roos (Eds.), *Processability approaches to language acquisition research & teaching* (Vol. 7, pp. 285–300). John Benjamins Publishing Company. https://doi.org/10.1075/palart.7.12roo

Roquet, H., & Pérez-Vidal, C. (2015). Do productive skills improve in content and language integrated learning contexts? The case of writing. *Applied Linguistics*, amv050. https://doi.org/10.1093/applin/amv050

Saeli, H., & Cheng, A. (2019). Effects of L1 writing experiences on L2 writing perceptions: evidence from an English as a foreign language context. *Reading & Writing Quarterly*, *35*(6), 509–524. https://doi.org/10.1080/10573569.2019.1579129

Sample, E., & Michel, M. (2015). An exploratory study into trade-off effects of complexity, accuracy, and fluency on young learners' oral task repetition. *TESL Canada Journal*, *31*, 23. https://doi.org/10.18806/tesl.v31i0.1185

Sato, M. (2017). Interaction mindsets, interactional behaviors, and L2 development: An affective-social-cognitive model. *Language Learning*, *67*(2), 249–283. https://doi.org/10.1111/lang.12214

Sato, M. (2020a). Metacognitive instruction for collaborative interaction: The process and product of self-regulated learning in the Chilean EFL context. In C. Lambert & R. Oliver (Eds.), *Using tasks in second language teaching: Practice in diverse contexts* (pp. 215–236). Multilingual Matters.

Sato, M. (2020b). Learner attitudes and attention to form in peer interaction: A proposal to replicate Adams et al. (2011) and Philp et al. (2010). *Language Teaching*, 1–10. https://doi.org/10.1017/S0261444820000610

Sato, M., & Ballinger, S. (Eds.). (2016). *Peer interaction and second language learning: Pedagogical potential and research agenda* (Vol. 45). John Benjamins Publishing Company. https://doi.org/10.1075/lllt.45

Sato, M., & Dussuel Lam, C. (2021). Metacognitive instruction with young learners: A case of willingness to communicate, L2 use, and metacognition of oral communication. *Language Teaching Research*, 13621688211004640. https://doi.org/10.1177/13621688211004639

Sato, M., & Loewen, S. (2018). Metacognitive Instruction enhances the effectiveness of corrective feedback: Variable effects of feedback types and linguistic targets: metacognitive instruction and corrective feedback. *Language Learning*, *68*(2), 507–545. https://doi.org/10.1111/lang.12283

Sato, M., & McDonough, K. (2020). Predicting L2 learners' noticing of L2 errors: Proficiency, language analytical ability, and interaction mindset. *System*, *93*, 102301. https://doi.org/10.1016/j.system.2020.102301

Sato, M., & Viveros, P. (2016). 3. Interaction or collaboration? Group dynamics in the foreign language classroom. In M. Sato & S. Ballinger (Eds.), *Language Learning & Language Teaching* (pp. 91–112). John Benjamins Publishing Company. https://doi.org/10.1075/lllt.45.04sat

Sato, M., McDonough, K., & Oyanedel, J. C. (2020). Data of the interaction mindset questionnaire: An initial exploration. *Data in Brief*. https://doi.org/10.1016/j.system.2020.102301

Schmidt, R. W. (1990). The role of consciousness in second language learning. *Applied Linguistics*, *11*(2), 129–158. https://doi.org/10.1093/applin/11.2.129

Schmidt, R. W. (2010). Attention, awareness, and individual differences in language learning. In W. M. Chan, S. Chi, K. N. Cin, J. Istanto, M. Nagami, J. W. Sew, T. Suthiwan, & I. Walker (Eds.), *Proceedings of CLaSIC 2010* (pp. 721–737). National University of Singapore, Centre for Language Studies.

Schoonen, R., Snellings, P., Stevenson, M., & Gelderen, A. van. (2009). Chapter 3. Towards a blueprint of the foreign language writer: The linguistic and cognitive demands of foreign language writing. In R. M. Manchón (Ed.), *Writing in foreign language contexts* (pp. 77–101). Multilingual Matters. https://doi.org/10.21832/9781847691859-007

Schoonen, R., van Gelderen, A., Stoel, R. D., Hulstijn, J., & de Glopper, K. (2011). Modeling the development of L1 and EFL writing proficiency of secondary school students: Modeling the development of L1 and EFL writing proficiency. *Language Learning*, *61*(1), 31–79. https://doi.org/10.1111/j.1467-9922.2010.00590.x

Schunk, D. H. (1984). Self-efficacy perspective on achievement behavior. *Educational Psychologist*, *19*(1), 48–58. https://doi.org/10.1080/00461528409529281

Seedhouse, P. (1999). Task-based interaction. *ELT Journal*, *53*(3), 149–156. https://doi.org/10.1093/elt/53.3.149

Serrano, R. (2011). The effect of program type and proficiency level in learners' written production. *Revista Española de Lingüística Aplicada (RESLA)*, *24*, 211–226.

Shak, J. (2006). Children using dictogloss to focus on form. *Reflections on English Language Teaching*, *5*(2), 47–62.

Shak, J., & Gardner, S. (2008). Young learner perspectives on four focus-on-form tasks. *Language Teaching Research*, *12*(3), 387–408. https://doi.org/10.1177/1362168808089923

Shehadeh, A. (2011). Effects and student perceptions of collaborative writing in L2. *Journal of Second Language Writing*, *20*(4), 286–305. https://doi.org/10.1016/j.jslw.2011.05.010

Shen, B., Bai, B., & Park, M. (2020). Exploring Hong Kong primary students' English writing motivation: Relationships between writing self-efficacy and task value. *Journal of Multilingual and Multicultural Development*, 1–15. https://doi.org/10.1080/01434632.2020.1823397

Shin, S.-Y., Lidster, R., Sabraw, S., & Yeager, R. (2016). The effects of L2 proficiency differences in pairs on idea units in a collaborative text reconstruction task. *Language Teaching Research*, *20*(3), 366–386. https://doi.org/10.1177/1362168814567888

Shintani, N. (2012). Input-based tasks and the acquisition of vocabulary and grammar: A process-product study. *Language Teaching Research*, *16*(2), 253–279. https://doi.org/10.1177/1362168811431378

Shintani, N. (2014). Using tasks with young beginner learners: The role of the teacher. *Innovation in Language Learning and Teaching*, *8*(3), 279–294. https://doi.org/10.1080/17501229.2013.861466

Shintani, N., Aubrey, S., & Donnellan, M. (2016). The effects of pre-task and post-task metalinguistic explanations on accuracy in second language writing. *TESOL Quarterly*, *50*(4), 945–955. https://doi.org/10.1002/tesq.323

Silva, R., & Clahsen, H. (2008). Morphologically complex words in L1 and L2 processing: Evidence from masked priming experiments in English. *Bilingualism: Language and Cognition*, *11*(2), 245–260. https://doi.org/10.1017/S1366728908003404

Silva, T. (1992). L1 vs L2 writing; ESL graduate students' perceptions. *TESL Canada Journal*, 27–47. https://doi.org/10.18806/tesl.v10i1.610

Sinclair, S., & Geoffrey, R. (2016). *Voyant Tools*. http://voyant-tools.org/

Skehan, P. (1998). *A cognitive approach to language learning*. Oxford Univ. Press.

Skehan, P. (2002). Theorising and updating aptitude. In P. Robinson (Ed.), *Individual differences and instructed language learning* (pp. 69–94). John Benjamins.

Skehan, P. (2004). *A cognitive approach to language learning*. Oxford Univ. Press.

Skehan, P. (2009). Modelling second language performance: Integrating complexity, accuracy, fluency, and lexis. *Applied Linguistics*, *30*(4), 510–532. https://doi.org/10.1093/applin/amp047

Skehan, P. (2018). *Second language task-based performance: Theory, research, assessment*. Routledge.

Skehan, P., & Foster, P. (1997). Task type and task processing conditions as influences on foreign language performance. *Language Teaching Research*, *1*(3), 185–211. https://doi.org/10.1177/136216889700100302

Skehan, P., & Foster, P. (2012). Complexity, accuracy, fluency and lexis in task-based performance: A synthesis of the Ealing research. In A. Housen, F. Kuiken, & I. Vedder (Eds.), *Language Learning & Language Teaching* (Vol. 32, pp. 199–220). John Benjamins Publishing Company. https://doi.org/10.1075/lllt.32.09fos

Skehan, P., & Shum, S. (2017). What influences performance? In L. L. C. Wong (Ed.), *Faces of English education* (1st ed., pp. 28–43). Routledge. https://doi.org/10.4324/9781315205618-3

Skinner, E., Furrer, C., Marchand, G., & Kindermann, T. (2008). Engagement and disaffection in the classroom: Part of a larger motivational dynamic? *Journal of Educational Psychology*, *100*(4), 765–781. https://doi.org/10.1037/a0012840

Soziolinguistika Klusterra. (2017). *Hizkuntzen erabileraren kale-neurketa. Euskal Herria, 2016.* (VII; p. 20). http://www.soziolinguistika.eus/files/hekn2016-_eu_1.pdf

Spada, N. (1997). Form-Focussed Instruction and Second Language Acquisition: A Review of Classroom and Laboratory Research. *Language Teaching*, *30*(2), 73–87. https://doi.org/10.1017/S0261444800012799

Spada, N., & Tomita, Y. (2010). Interactions between type of instruction and type of language feature: A meta-analysis: Type of instruction and language feature. *Language Learning*, *60*(2), 263–308. https://doi.org/10.1111/j.1467-9922.2010.00562.x

Stauble, A. M. (1984). A comparison of a Spanish-English and a Japanese-English second language continuum: Negation and verb morphology. In R. W. Andersen (Ed.), *Second Languages: A cross-linguistic perspective* (pp. 323–353). Newbury House.

Stevens, J. (1996). *Applied multivariate statistics for the social sciences*. Lawrence Erlbaum Associates. https://books.google.es/books?id=08EfAQAAIAAJ

Storch, N. (1997). The editing talk of adult ESL learners. *Language Awareness*, *6*(4), 221–232. https://doi.org/10.1080/09658416.1997.9959931

Storch, N. (1999). Are two heads better than one? Pair work and grammatical accuracy. *System*, *27*(3), 363–374. https://doi.org/10.1016/S0346-251X(99)00031-7

Storch, N. (2001). Comparing ESL learners' attention to grammar on three different classroom tasks. *RELC Journal*, *32*(2), 104–124. https://doi.org/10.1177/003368820103200207

Storch, N. (2002). Patterns of interaction in ESL pair work. *Language Learning*, *52*(1), 119–158. https://doi.org/10.1111/1467-9922.00179

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*(2), 95–114. https://doi.org/10.1080/09658410802146644

Storch, N. (2009a). *The nature of pair interaction: Learners' interaction in an ESL class ; its nature and impact on grammatical development*. VDM Verl. Müller.

Storch, N. (2009b). The impact of studying in a second language (L2) medium university on the development of L2 writing. *Journal of Second Language Writing*, *18*(2), 103–118. https://doi.org/10.1016/j.jslw.2009.02.003

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

Storch, N. (2016). Collaborative writing. In R. M. Manchón & P. K. Matsuda (Eds.), *Handbook of second and foreign language writing* (pp. 387–406). De Gruyter. https://doi.org/10.1515/9781614511335-021

Storch, N. (2017). Sociocultural theory in the L2 classroom. In S. Loewen & M. Sato (Eds.), *The Routledge handbook of instructed second language acquisition* (pp. 69–83). Routledge. https://doi.org/10.4324/9781315676968-5

Storch, N. (2018). Collaborative writing TESOL entry. In J. I. Liontas, T. International Association, & M. DelliCarpini (Eds.), *The TESOL encyclopedia of English language teaching* (pp. 1–6). John Wiley & Sons, Inc. https://doi.org/10.1002/9781118784235.eelt0395

Storch, N. (2019a). Collaborative writing. *Language Teaching*, *52*(01), 40–59. https://doi.org/10.1017/S0261444818000320

Storch, N. (2019b). Collaborative writing as peer feedback. In K. Hyland & F. Hyland (Eds.), *Feedback in second language writing* (pp. 143–162). Cambridge University Press. https://doi.org/10.1017/9781108635547.010

Storch, N., & Aldosari, A. (2010). Learners' use of first language (Arabic) in pair work in an EFL class. *Language Teaching Research*, *14*(4), 355–375. https://doi.org/10.1177/1362168810375362

Storch, N., & Aldosari, A. (2013). Pairing learners in pair work activity. *Language Teaching Research*, *17*(1), 31–48. https://doi.org/10.1177/1362168812457530

Storch, N., & Wigglesworth, G. (2010). Students' engagement with feedback on writing: The role of learner agency/beliefs. In R. Batstone (Ed.), *Sociocognitive perspectives on language use and language learning* (pp. 166–185). Oxford University Press.

Ströbel, M., Kerz, E., & Wiechmann, D. (2020). The relationship between first and second language writing: Investigating the effects of first language complexity on second language complexity in advanced stages of learning: L1-L2 writing relationship. *Language Learning*. https://doi.org/10.1111/lang.12394

Strutt, G. F., Anderson, D. R., & Well, A. D. (1975). A developmental study of the effects of irrelevant information on speeded classification. *Journal of Experimental Child Psychology*, *20*(1), 127–135. https://doi.org/10.1016/0022-0965(75)90032-6

Suzuki, W., & Storch, N. (Eds.). (2020). *Languaging in language learning and teaching: A collection of empirical studies* (Vol. 55). John Benjamins Publishing Company. https://doi.org/10.1075/lllt.55

Suzuki, Y., & DeKeyser, R. (2017). The interface of explicit and implicit knowledge in a second language: Insights from individual differences in cognitive aptitudes: interface of explicit and implicit knowledge. *Language Learning*, *67*(4), 747–790. https://doi.org/10.1111/lang.12241

Svalberg, A. M.-L. (2018). Researching language engagement; current trends and future directions. *Language Awareness*, *27*(1–2), 21–39. https://doi.org/10.1080/09658416.2017.1406490

Svalberg, A. M.-L., & Askham, J. (2020). Teacher and learner perceptions of adult foreign language learners' engagement with consciousness-raising tasks in four languages. *Language Awareness*, 1–19. https://doi.org/10.1080/09658416.2020.1785483

Swain, M. (1985). Communicative competence: Some roles of comprehensible input and comprehensible output in its development. In S. Gass & M. Madden (Eds.), *Input in second language acquisition* (pp. 235–253). Newbury House.

Swain, M. (1995). Three functions of output in second language learning. In G. Cook & B. Seidlhofer (Eds.), *Principle and practice in applied linguistics: Studies in honour of H. G. Widdowson* (pp. 125–144). Oxford University Press.

Swain, M. (1998). Focus on form through conscious reflection. In C. Doughty & J. Williams (Eds.), *Focus on form in classroom second language acquisition* (pp. 64–81). Cambridge University Press.

Swain, M. (2000). The output hypothesis and beyond: Mediating acquisition through collaborative dialogue. In J. Lantolf (Ed.), *Sociocultural Theory and Second Language Learning* (pp. 97–114). Oxford University Press.

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

Swain, M. (2006). Languaging, agency and collaboration in advanced second language proficiency. In H. Byrnes (Ed.), *Advanced language learning: The contribution of Halliday and Vygotsky* (pp. 95–108). Continuum.

Swain, M. (2010). Talking it through: Languaging as a source of learning. In R. Batstone (Ed.), *Socicognitive perspectives on language use and language learning* (pp. 112–130). Oxford University Press.

Swain, M., & Lapkin, S. (1991). Additive bilingualism and French immersion education: The roles of language proficiency and literacy. In *Bilingualism, multiculturalism, and second language learning: The McGill conference in honour of Wallace E. Lambert.* (pp. 203–216). Lawrence Erlbaum Associates, Inc.

Swain, M., & Lapkin, S. (1995). Problems in output and the cognitive processes they generate: A step towards second language learning. *Applied Linguistics*, *16*(3), 371–391. https://doi.org/10.1093/applin/16.3.371

Swain, M., & Lapkin, S. (1998). Interaction and second language learning: Two adolescent French immersion students working together. *The Modern Language Journal*, *82*(3), 320–337. https://doi.org/10.1111/j.1540-4781.1998.tb01209.x

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). Pearson Education Limited. https://doi.org/10.4324/9781315838267-14

Swain, M., Lapkin, S., Knouzi, I., Suzuki, W., & Brooks, L. (2009). Languaging: University students learn the grammatical concept of voice in French. *The Modern Language Journal*, *93*(1), 5–29. https://doi.org/10.1111/j.1540-4781.2009.00825.x

Tagarelli, K. M., Ruiz, S., Vega, J. L. M., & Rebuschat, P. (2016). Variability in second language learning: The roles of individual differences, learning conditions, and linguistic complexity. *Studies in Second Language Acquisition*, *38*(2), 293–316. https://doi.org/10.1017/S0272263116000036

Tavakoli, P., & Skehan, P. (2005). 9. Strategic planning, task structure and performance testing. In R. Ellis (Ed.), *Planning and Task Performance in a Second Language* (pp. 239–273). John Benjamins Publishing Company. https://doi.org/10.1075/lllt.11.15tav

Tejada-Sánchez, I., & Pérez-Vidal, C. (2018). Writing performance and time of exposure in EFL immersion learners: Analysing complexity, accuracy, and fluency. In C. Pérez-Vidal, S. López-Serrano, J. Ament, & D. J. Thomas-Wilhelm (Eds.), *Learning context effects: Study abroad, formal instruction and international immersion classrooms* (pp. 101–129). Language Science Press. https://zenodo.org/record/1446470

Tellier, A. (2015). *Metalinguistic awareness and foreign language learning in primary schools: A classroom study with children aged 8 to 9 years*. University of Essex (unpublished).

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

Terrazas Gallego, M., & Agustín Llach, M. P. (2009). Exploring the increase of receptive vocabulary knowledge in the foreign language: A longitudinal study. *International Journal of English Studies*, *9*(113–133).

The United Nations. (1989). Convention on the rights of the child. *Treaty Series*, *1577*, 3.

Torras, M. R. (2005). Procesos psicolingüísticos implicados en la adquisición del inglés en el contexto de la enseñanza primaria. *Lenguaje y textos*, *23*, 89–112.

Torras, M. R., Navés, T., Celaya, M. L., & Pérez-Vidal, C. (2006). Age and IL development in writing. In C. Muñoz (Ed.), *Age and rate of foreign language learning* (pp. 156–182). Multilingual Matters.

Tragant, E., Marsol, A., Serrano, R., & Llanes, À. (2016). Vocabulary learning at primary school: A comparison of EFL and CLIL. *International Journal of Bilingual Education and Bilingualism*, *19*(5), 579–591. https://doi.org/10.1080/13670050.2015.1035227

Tragant, E., Thompson, M. S., & Victori, M. (2013). Understanding foreign language learning strategies: A validation study. *System, 41*(1), 95–108. https://doi.org/10.1016/j.system.2013.01.007

Treffers-Daller, J., Parslow, P., & Williams, S. (2016). Back to basics: How measures of lexical diversity can help discriminate between CEFR levels. *Applied Linguistics*, amw009. https://doi.org/10.1093/applin/amw009

Uggen, M. S. (2012). Reinvestigating the noticing function of output: Noticing function of output. *Language Learning*, *62*(2), 506–540. https://doi.org/10.1111/j.1467-9922.2012.00693.x

Ushioda, E. (2012). Motivation: L2 Learning as a Special Case? In S. Mercer, S. Ryan, & M. Williams (Eds.), *Psychology for Language Learning: Insights from Research, Theory and Practice* (pp. 58–73). Palgrave Macmillan UK. https://doi.org/10.1057/9781137032829_5

VanPatten, B., & Williams, J. (2014). Introduction. In B. VanPatten & J. Williams (Eds.), *Theories in second language acquisition: An introduction* (pp. 1–16). Routledge.

Villarreal, I., & Gil-Sarratea, N. (2019). The effect of collaborative writing in an EFL secondary setting. *Language Teaching Research*, 136216881982901. https://doi.org/10.1177/1362168819829017

Villarreal, I., & Munarriz-Ibarrola, M. (2021). "Together we do better": The effect of pair and group work on young EFL learners' written texts and attitudes. In M. P. García Mayo (Ed.) *Working Collaboratively in second/foreign language learning* (pp. 89–115). De Gruyter.

Vraciu, A., & Pladevall-Ballester, E. (2020). L1 use in peer interaction: Exploring time and proficiency pairing effects in primary school EFL. *International Journal of Bilingual Education and Bilingualism*, 1–18. https://doi.org/10.1080/13670050.2020.1767029

Vyatkina, N. (2012). The development of second language writing complexity in groups and individuals: A longitudinal learner corpus study: The development of second language writing complexity in groups and individuals. *The Modern Language Journal*, *96*(4), 576–598. https://doi.org/10.1111/j.1540-4781.2012.01401.x

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

Wajnryb, R. (1990). *Grammar dictation*. OUP Oxford.

Wang, Q., & Li, S. (2019). The relationship between task motivation and L2 motivation: An empirical study. In Z. Wen & M. J. Ahmadian (Eds.), *Researching L2 task performance and pedagogy: In honour of Peter Skehan* (pp. 67–92). John Benjamins Publishing Company. https://doi.org/10.1075/tblt.13.05wan

Watanabe, Y. (2019). 4. The role of languaging in collaborative and individual writing: when pairs outperform individuals. In M. Haneda & H. Nassaji (Eds.), *Perspectives on language as action* (pp. 63–79). Multilingual Matters. https://doi.org/10.21832/9781788922944-007

Watanabe, Y., & Swain, M. (2007). Effects of proficiency differences and patterns of pair interaction on second language learning: Collaborative dialogue between adult ESL learners. *Language Teaching Research*, *11*(2), 121–142. https://doi.org/10.1177/136216880607074599

Weigle, S. C. (2012). Assessment of Writing. In C. A. Chapelle (Ed.), *The encyclopedia of applied linguistics*. Blackwell Publishing Ltd. https://doi.org/10.1002/9781405198431.wbeal0056

Wen, Z. (Edward), & Ahmadian, M. J. (Eds.). (2019). *Researching L2 task performance and pedagogy: In honour of Peter Skehan* (Vol. 13). John Benjamins Publishing Company. https://doi.org/10.1075/tblt.13

White, J. (2008). Speeding up acquisition of "his" and "her": Explicit L1/L2 contrasts help. In J. Philp, R. Oliver, & A. Mackey (Eds.), *Second language acquisition and the younger learner: Child's play?* (pp. 193–228). John Benjamins Publishing Company. https://doi.org/10.1075/lllt.23.13whi

White, J., Muñoz, C., & Collins, L. (2007). The his her challenge: Making progress in a 'regular' L2 programme. *Language Awareness*, *16*(4), 278–299. https://doi.org/10.2167/la419.0

White, L., Spada, N., Lightbown, P. M., & Ranta, L. (1991). Input enhancement and L2 question formation. *Applied Linguistics*, *12*(4), 416–432. https://doi.org/10.1093/applin/12.4.416

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

Wigglesworth, G., & Storch, N. (2012). What role for collaboration in writing and writing feedback. *Journal of Second Language Writing*, *21*(4), 364–374. https://doi.org/10.1016/j.jslw.2012.09.005

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

Willis, D., & Willis, J. R. (2011). *Doing task-based teaching* (5. print). Oxford University Press.

Willis, J. (1996). *A framework for task-based learning*. Longman.

Wolfe-Quintero, K., Inagaki, S., Kim, H. Y., & Center, U. of H. at M. S. L. T. & C. (1998). *Second language development in writing: Measures of fluency, accuracy, & complexity*. Second Language Teaching & Curriculum Center, University of Hawaii at Manoa. https://books.google.es/books?id=IboEPPjPGgkC

Wood, D. (1998). *How children think and learn: The social contexts of cognitive development* (2nd ed). Blackwell.

Wood, D., Bruner, J. S., & Ross, G. (1976). The role of tutoring in Problem Solving. *Journal of Child Psychology and Psychiatry*, *17*(2), 89–100. https://doi.org/10.1111/j.1469-7610.1976.tb00381.x

Yamashita, J. (2001). Transfer of L1 reading ability to L2 reading: An elaboration of the linguistic threshold. *Studies in Language and Culture*, *23*(1), 189–200.

Yan, C. M. W., McBride-Chang, C., Wagner, R. K., Zhang, J., Wong, A. M. Y., & Shu, H. (2012). Writing quality in Chinese children: Speed and fluency matter. *Reading and Writing*, *25*(7), 1499–1521. https://doi.org/10.1007/s11145-011-9330-y

Yang, L., & Zhang, L. (2010). Exploring the role of reformulations and a model text in EFL students' writing performance. *Language Teaching Research, 14*(4), 464–484. https://doi.org/10.1177/1362168810375369

Yu, S., & Lee, I. (2016). Peer feedback in second language writing (2005–2014). *Language Teaching*, *49*(04), 461–493. https://doi.org/10.1017/S0261444816000161

Zalbidea, J. (2021). On the scope of output in SLA: Task modality, salience, L2 grammar noticing, and development. *Studies in Second Language Acquisition*, *43*(1), 50–82. https://doi.org/10.1017/S0272263120000261

Zhang, M. (2019a). Understanding L1 and L2 interaction in collaborative writing: A lexico-grammatical analysis. *Language Teaching Research*, 136216881985991. https://doi.org/10.1177/1362168819859911

Zhang, M. (2019b). Towards a quantitative model of understanding the dynamics of collaboration in collaborative writing. *Journal of Second Language Writing*, *45*, 16–30. https://doi.org/10.1016/j.jslw.2019.04.001

Zhang, M., & Plonsky, L. (2020). Collaborative writing in face-to-face settings: A substantive and methodological review. *Journal of Second Language Writing*, *49*, 100753. https://doi.org/10.1016/j.jslw.2020.100753

RESUMEN EN ESPAÑOL

Antecedentes

Existe una gran cantidad de evidencia empírica que constata que la interacción entre pares constituye una actividad fundamental para el aprendizaje de una segunda lengua (L2) (Philp et al., 2014; Sato & Ballinger, 2016). La investigación llevada a cabo en este campo se apoya principalmente sobre dos marcos teóricos complementarios: el cognitivo-interaccionista y la teoría sociocultural. Según el cognitivo-interaccionista (Long, 1983, 1996), cuando los aprendices de una L2 interaccionan, tienen que resolver las faltas de entendimiento por medio de la negociación de significado. En este proceso, reciben input modificado al tiempo que también modifican su propia producción (Swain, 2005), lo que les lleva a activar ciertos mecanismos cognitivos, como el reconocimiento *(noticing,* Schmidt, 1990, 2010) de las diferencias entre su conocimiento de la L2 y lo que son capaces de expresar en la misma *(noticing-the-gap)*. Por otra parte, de acuerdo con la teoría sociocultural (Vygotsky, 1978), el aprendizaje se enmarca en una determinada situación social, y los aprendices se valen de diversos mediadores para alcanzar procesos psicológicos de orden superior. Entre ellos, el uso del lenguaje destaca como el más importante. Trasladado al contexto de aprendizaje de una L2 (Lantolf, 2000; Ohta, 2001), cuando los aprendices colaboran en una tarea, se proporcionan ayuda (andamiaje) y co-construyen conocimiento que luego internalizan en su conocimiento individual. Este *diálogo colaborativo* (Swain, 2000) se considera no tanto un proceso facilitador del aprendizaje sino un aprendizaje en sí mismo, y en múltiples ocasiones trata en torno a cuestiones lingüísticas. Dichos casos se suelen operacionalizar mediante el concepto de Episodios Relacionados con la Lengua (ERL) (Swain & Lapkin, 1998).

La interacción entre pares ha cobrado si cabe más importancia gracias al enfoque comunicativo (Breen & Candlin, 1980), del que emana el marco de aprendizaje de lenguas basado en tareas (TBLT, en sus siglas en inglés) (R. Ellis et al., 2020). Los estudios realizados en este campo han tratado de entender qué tareas pueden activar los procesos de adquisición de una L2 descritos anteriormente, así como bajo qué circunstancias (relacionadas con el diseño de la propia tarea o el aprendiz) resultan más beneficiosas. En este sentido, la investigación se ha llevado a cabo tradicionalmente con adultos, mientras que los jóvenes aprendices de lenguas extranjeras en contextos formales no han recibido tanta atención (García Mayo, 2018b). Sin embargo, esta tendencia se está invirtiendo en los últimos años, gracias principalmente al incremento de programas de aprendizaje de inglés como lengua extranjera (ILE) en Educación Primaria (Enever, 2018). Así pues, cabe señalar que los resultados obtenidos con adultos no son extrapolables a los de la población infantil, ya que existen notables diferencias entre ellos, incluidos los mecanismos de aprendizaje, la capacidad de atención, la memoria, el grado de conocimiento de la primera lengua y las necesidades afectivas (Azkarai & Oliver, 2018; Butler, 2016; Roehr-Brackin & Tellier, 2019).

Teniendo en consideración todo lo anterior, la presente tesis pretende arrojar luz sobre la potencialidad de un tipo de tarea de escritura colaborativa, más concretamente, la tarea de la dictoglosia (Wajnryb, 1990), para estimular la atención a la forma (Long, 1991) y el desarrollo de la escritura en L2 en aprendices de ILE del último curso de Educación Primaria. Las tareas de escritura colaborativa (Storch, 2016, 2019a) se caracterizan por aunar los beneficios de la interacción oral y las ventajas de la modalidad escrita. De hecho, pese haber sido relegada a un segundo plano en el enfoque de TBLT (Byrnes & Manchón, 2014), se ha demostrado que la escritura posee ciertas características que la convierten en un contexto idóneo para el aprendizaje de la L2 (Manchón & Vasylets, 2019). En concreto, cabe destacar su ritmo más pausado y el registro visual de la producción, lo que permite una monitorización más exhaustiva de la producción. Desde un punto de vista psicolingüístico, esto supone la posibilidad de recurrir a la memoria de largo plazo y de detectar problemas en el output (Gilabert et al., 2016; Zalbidea, 2021).

Los estudios acerca de las tareas de escritura colaborativa se han centrado en dos principales líneas de investigación (Storch, 2016, 2019b): por un lado, el análisis de los factores (bien relacionados con las tareas o con el aprendiz) que influyen sobre la ocurrencia de ERL y, por otro, el estudio del efecto de ese diálogo colaborativo en el producto escrito, así como en la adquisición de la L2. Por ello, en el estudio presentado en esta tesis, aparte de examinar una población infrarrepresentada, nos proponemos adoptar estas dos líneas de investigación analizando el efecto de dos variables que por ahora han recibido escasa atención en este campo (Zhang & Plonsky, 2020), en concreto, la repetición de la tarea (RT) (Bygate, 2018) y la provisión de una instrucción focalizada hacia la forma (IFF) previa a la tarea (DeKeyser, 1998). Paralelamente, dada la importancia de la dimensión afectiva en la eficacia de las tareas comunicativas (Lambert, 2017) y en el aprendizaje temprano de lenguas en general (Mihaljević Djigunović & Nikolov, 2019), pretendemos obtener un mayor conocimiento sobre la predisposición de los jóvenes aprendices hacia la escritura y el trabajo colaborativo, y acerca de su visión sobre la tarea de la dictoglosia. Por último, dado que hay voces que señalan que los logros en escritura en L2 dependen en cierta medida de las actitudes y la competencia escrita en L1 de los aprendices (Pae, 2008, 2018; Schoonen et al., 2011), tratamos de probar la relación entre estas tres variables.

El estudio

<u>Objetivo y preguntas de investigación</u>

El estudio que presentamos en esta tesis doctoral tiene como objetivo determinar la eficacia de la tarea de la dictoglosia para orientar la atención de los jóvenes aprendices de ILE (11-12 años) hacia la forma, así como entender el impacto del trabajo colaborativo en la calidad del producto escrito de la tarea y en el desarrollo de la escritura en L2. Asimismo, intentamos comprobar la influencia de la RT y la provisión de IFF previa a la tarea tanto sobre la atención a la forma como en la calidad de la escritura. Por último, queremos analizar ciertas diferencias individuales, como las actitudes de estos alumnos hacia la escritura y el trabajo colaborativo o hacia la propia tarea, y obtener evidencia sobre la posibilidad de predecir los logros en escritura en L2 mediante la competencia escrita en L1 y las actitudes hacia la escritura. Con estos objetivos en mente, formulamos las siguientes preguntas de investigación, agrupadas en tres grandes bloques temáticos:

- RQ1: Dictoglosia, atención a la forma y el impacto de variables relacionadas con la tarea
  - RQ1a: ¿Cuál es el foco de los Episodios Relacionados con la Lengua (ERL), tanto escritos como orales?
  - RQ1b: En cuanto a los ERL orales, ¿cómo los resuelven?, ¿qué nivel de implicación *(engagement)* demuestran? ¿qué uso realizan de la L2 y las lenguas previas (LPs)?

    - RQ1c: ¿Qué influencia ejerce la repetición del procedimiento de la tarea (RT) y la provisión de instrucción focalizada hacia la forma (IFF) sobre la atención a la forma?
    - RQ1d: ¿Se centran los jóvenes aprendices en las formas meta (3S y POSS)?
- RQ2: Trabajo colaborativo
    - RQ2a: ¿Qué patrones de interacción manifiestan los jóvenes aprendices durante la dictoglosia colaborativa? ¿Existe una influencia de la RT?
    - RQ2b: ¿En qué medida influyen el trabajo colaborativo, la RT y la IFF en la calidad de la reconstrucción del texto de la dictoglosia?
    - RQ2c: ¿Tiene algún efecto la tarea de la dictoglosia en la escritura individual en L2?
- RQ3: Diferencias individuales
    - RQ3a: ¿Qué actitudes manifiestan los jóvenes aprendices hacia la escritura y el trabajo colaborativo?
    - RQ3b: ¿Qué percepción tienen sobre la tarea de la dictoglosia?
    - RQ3c: ¿Cuál es el impacto de la competencia escrita en L1 y las actitudes hacia la escritura sobre los logros en la escritura en L2?

Metodología

Participantes

En este estudio participaron 93 jóvenes aprendices de ILE, de edades entre 11-12 años, que cursaban 6º de Educación Primaria en dos colegios públicos distintos de Vitoria-Gasteiz. Todos ellos tenían el español como L1 y asistían a un modelo lingüístico de inmersión en euskera (modelo D). Por otra parte, recibían 3 horas de ILE en el colegio y habían estudiado este idioma una media cercana a 7 años. Asimismo, un porcentaje reseñable (43,3 %) afirmaba asistir a clases extracurriculares de inglés. Tal y como pudimos constatar en la prueba de nivel *Flyers* (Cambridge Assessment English, 2018) administrada con anterioridad al procedimiento experimental, la proficiencia promedia en inglés era de un nivel A1 (básico) (Council of Europe, 2001, 2018).

Tarea y procedimiento

La tarea elegida para este estudio fue la tarea de la dictoglosia (Wajnryb, 1990), que consiste en escuchar dos veces un breve texto y reconstruirlo de manera correcta y lo más parecida posible. Para ello, se diseñaron dos textos que tenían dos formas meta distintas: la tercera persona del singular *-s* del presente (3S) y los determinantes posesivos de tercera persona *his/her* (POSS). Anteriores estudios han señalado que estas formas difieren de forma sustancial tanto en su prominencia cognitiva como en su valor comunicativo (Sato & Loewen, 2018). Asimismo, se ha demostrado que la 3S es de las últimas formas en ser adquirida en inglés como L2, y además resulta problemática para los aprendices ILE que tienen español como L1 (Dulay & Burt, 1974). Por otra parte, se ha constatado que POSS resulta especialmente compleja para los hablantes de español como L1 (J. White, 2008), así como para los bilingües español-euskera (Imaz Agirre & García Mayo, 2013, 2018).

Los participantes realizaron la tarea en dos días distintos en el intervalo de dos semanas consecutivas. El orden de los textos fue aleatorio para evitar una posible influencia de la forma meta. Asignamos a los aprendices a tres condiciones diferentes: aquellos que llevaron a cabo el procedimiento canónico de dictoglosia en parejas (Colab), los que recibieron una IFF y, tras ello, realizaron la dictoglosia por parejas (IFF+Collab) y, por último, los que la realizaron de manera individual, también referido como grupo de comparación (Comp). Dicha asignación vino en parte condicionada por el colegio al que asistían los participantes, por tanto, consideramos que el diseño de este estudio es cuasiexperimental (Loewen & Plonsky, 2015). En lo que se refiere a la IFF, diseñamos dos procedimientos, uno dirigido a la 3S y otro a POSS, en los que los aprendices recibían información explícita sobre la forma y llevaban a cabo ejercicios de comprensión y producción. Por otra parte, los participantes de Comp recibían una ficha al terminar de reconstruir el texto de la dictoglosia con el fin de que pudieran reflexionar individualmente por escrito sobre cuestiones metalingüísticas (Hanaoka & Izumi, 2012; Ishikawa, 2018; Ishikawa & Révész, 2020).

Por último, antes y después del procedimiento de la dictoglosia llevaron a cabo pruebas de escritura narrativa en L2 para comprobar los efectos de la tarea en su capacidad de producción escrita individual. En concreto realizaron un pretest (T1) una semana antes de la primera dictoglosia, un postest (T2) y un postest diferido (T3), una y dos semanas más tarde de la última dictoglosia, respectivamente. Asimismo, antes del procedimiento, cumplimentaron un cuestionario sobre sus actitudes hacia la escritura y el trabajo colaborativo, y una vez realizada la tarea, completaron otro sobre su percepción de la dictoglosia. Su visión acerca de la tarea fue además complementada por entrevistas focales realizadas al término del procedimiento experimental. Finalmente, se les proporcionó un ejercicio de escritura narrativa en L1 para comprobar su aptitud en la producción escrita en español.

Análisis de los datos

La interacción oral de los aprendices durante la tarea de la dictoglosia fue transcrita y analizada en términos de cantidad y calidad de ERL (García Mayo & Azkarai, 2016; McDonough & Hernández González, 2019; Storch, 2008) mediante el programa NVivo (QSR International, 2012). El uso predominante de las LP y la L2 en estos episodios se analizó por medio de CLAN (MacWhinney, 2000), mientras que las dinámicas de pareja se clasificaron holísticamente atendiendo a los criterios de equidad y mutualidad (Storch, 2002).

En lo que se refiere a la producción escrita en L2, esta fue transcrita a Excel y fue analizada de forma manual. Para la complejidad y la corrección, se emplearon medidas textuales tradicionales (p. ej. cláusulas por unidad T o número de errores por 100 palabras), así como otras más específicas de la tarea (p. ej. corrección de las formas meta 3S y POSS o cantidad de unidades idea recuperadas del texto original, Carrell, 1985). Paralelamente, también se evaluó a través de una rúbrica analítica. En el caso de la escritura en L1, su calidad se midió en función de la escala de una rúbrica ya validada (Fernández et al., 2019). Respecto a los cuestionarios, las respuestas fueron trascritas a Excel y fueron objeto de un minucioso proceso de validación. Las preguntas abiertas y las entrevistas focales, por su parte, fueron analizadas en función temas y términos clave con el programa NVivo (QSR International, 2012).

Los datos cuantitativos se analizaron con los *software* estadísticos SPSS (IBM Corp, 2016) y JASP (JASP Team, 2019). Dependiendo de las preguntas de investigación, se recurrió a diversos análisis inferenciales, como la ANOVA de medidas repetidas, la regresión múltiple o el análisis factorial, todos ellos comunes al campo de la lingüística aplicada (Larson-Hall, 2016).

Resultados

RQ1

La mayoría de los ERL que generaron los jóvenes aprendices durante la tarea colaborativa estuvieron relacionados con aspectos mecánicos (ortografía y puntuación) y otras formas gramaticales (es decir, distintas a 3S y POSS). Sin embargo, los episodios más largos fueron los relacionados con el léxico y otras formas gramaticales. Cabe destacar que la implicación por parte de estos participantes en las discusiones metalingüísticas fue, en general, limitada, con la excepción de los episodios léxicos, en los que predominó una implicación elaborada. Asimismo, en lo que al uso de lenguas se refiere, hubo un uso mayoritario de las LP para la resolución de los episodios, siendo aún más notable en el caso de los episodios léxicos y mecánicos. Respecto al tipo de resolución, hay que señalar que, independientemente del foco, los jóvenes aprendices fueron capaces de resolver correctamente la mayoría de los ERL.

La influencia de la variable de RT fue mínima tanto en la cantidad como el tipo de atención a la forma. Aun así, el grupo IFF+Colab redujo de manera significativa el tiempo empleado en la tarea del primer al segundo día. La atención a las formas meta (3S y POSS) fue en general baja. Con todo, hubo significativamente más episodios sobre 3S cuando esta forma era el objetivo del texto de la dictoglosia que cuando no lo era (y viceversa en el caso de POSS). Además, añadir una IFF antes de la tarea generó un promedio significativamente mayor de episodios sobre 3S que en el contexto de la dictoglosia canónica (Colab). Sin embargo, esta ventaja de IFF+Colab sobre Colab no se dio en el caso de POSS.

Finalmente, el bajo número de comentarios metalingüísticos escritos evidenciaron que los aprendices que realizaron la tarea de forma individual (Comp) tuvieron serias dificultades para llevar a cabo esta tarea de reflexión sobre la lengua, y que cuando así lo hicieron estos episodios trataron sobre aspectos léxicos y mecánicos.

RQ2

Las dinámicas de pareja que manifestaron los jóvenes aprendices durante la tarea de la dictoglosia fueron principalmente colaborativas, aunque en el caso de IFF+Colab la proporción de patrones no colaborativos fue mayor que en Colab. La RT no tuvo un gran impacto sobre estos comportamientos, y los cambios entre el primer y el segundo día no implicaron necesariamente más patrones colaborativos, sino que en determinadas ocasiones los aprendices demostraron dinámicas más cooperativas o dominantes/pasivas.

Respecto al impacto del trabajo colaborativo en la calidad de la reconstrucción, no hallamos evidencia de una ventaja de la escritura en parejas sobre la individual. No obstante, se pudo observar una tendencia en Colab y, especialmente, en IFF+Colab, a aumentar la corrección gramatical del primer al segundo día, mientras que dichas medidas en el caso de Comp no se vieron alteradas. Por otra parte, la RT no surtió efectos significativos en las medidas de escritura, si bien los aprendices, independientemente de su condición en la tarea, utilizaron significativamente menos coordinación durante la segunda reconstrucción. Por último, aunque no de manera significativa, IFF+Colab fue superior a los otros dos grupos en cuanto a la corrección de las formas meta.

Para terminar, en referencia a la influencia de la dictoglosia en la producción escrita individual en L2, los resultados no permitieron observar un aumento de la calidad en términos generales. El único grupo donde se evidenció una mejoría significativa de la corrección gramatical fue IFF+Colab, que experimentó un aumento progresivo de cláusulas gramaticales y contextualizadas, lo que finalmente se tradujo en una diferencia significativa del T1 al T3. El tipo de cláusulas también indicó que los aprendices, independientemente de su grupo, generaron significativamente menos precláusulas (cláusulas agramaticales y descontextualizadas) en el T2 respecto al T1. Otras medidas de corrección gramatical (como el número de errores por cada 100 palabras) también apuntaron, aunque no de forma significativa, a una mejoría de esta dimensión del T1 al T2 en el caso de IFF+Colab.

RQ3

Los aprendices demostraron una predisposición favorable tanto a colaborar como a escribir, si bien en el grupo IFF+Colab esta última resultó ser significativamente menor que en Colab y Comp. Por otra parte, las respuestas abiertas demostraron que desde un principio todos eran más reticentes a escribir en L2 que en L1, y de nuevo, esta tendencia fue más acusada en el caso de IFF+Colab que en el resto.

En cuanto a la percepción de la tarea de la dictoglosia, los participantes que la realizaron de forma colaborativa manifestaron una actitud más positiva que aquellos que la completaron de manera individual, si bien los niveles de aceptación fueron en general bastante altos. Los aprendices de ambos modos de trabajo coincidieron en expresar un menor entusiasmo por la parte de la reescritura, y, por el contrario, una preferencia hacia la escucha de los audios. Asimismo, no consideraron la tarea excesivamente compleja y percibieron que repetirla conllevaba beneficios. Aun así, en las entrevistas algunos alumnos afirmaron que llevar a cabo la dictoglosia una segunda vez en las mismas condiciones (por ejemplo, en las mismas parejas) les resultó desmotivador. Cabe mencionar que los participantes percibieron que el proceso experimental en conjunto había resultado beneficioso, ya que les había permitido trabajar la L2 de forma distinta a sus rutinas de clase habituales y comprender mejor cuál era su nivel en este idioma.

Por último, en lo que se refiere a la capacidad de predicción de la competencia escrita en L1 y de las actitudes hacia la escritura sobre la calidad de la escritura en L2, el modelo de regresión estimó una mayor contribución de la variable afectiva (14 %) que de la cognitiva (6 %), mientras que el conjunto de las dos variables predictoras alcanzaba a explicar un 27 % de la varianza total en la calidad de la escritura en L2.

Conclusiones

El presente estudio ha servido para esclarecer los efectos de la tarea de la dictoglosia en el aula de ILE del último curso de Educación Primaria. En línea con anteriores trabajos que analizaron el impacto de esta tarea en la atención a la forma de esta población (Calzada & García Mayo, 2020a, 2021a), los resultados de la presente tesis confirman que constituye una herramienta adecuada para generar episodios sobre aspectos formales de la lengua (mecánicos y relativos a formas gramaticales distintas a las meta). Así pues, se corrobora el papel del componente escrito para atraer más atención hacia la forma lingüística que hacia el significado (García Mayo & Azkarai, 2016).

A pesar de su bajo nivel en inglés, estos jóvenes aprendices fueron capaces de resolver correctamente la gran mayoría de discusiones metalingüísticas. Por otra parte, observamos que, al contrario de lo que sucede en tareas exclusivamente orales (Azkarai & García Mayo, 2017), la modalidad escrita incrementa el uso de lenguas previas frente a la L2 (Martínez Adrián & Arratibel Irazusta, 2020), indicando que su uso se asimila al de una herramienta cognitiva (Vraciu & Pladevall-Ballester, 2020). En este sentido, cabe destacar que los episodios más largos y donde hubo una implicación más elaborada de los participantes (en concreto, los ERL léxicos) fue donde el uso de LP resultó muy extendido.

Las dos variables relacionadas con la tarea, la repetición de la tarea y la provisión de IFF, no demostraron ejercer una gran influencia sobre los ERL. Por una parte, los aprendices generaron un número parecido de episodios ambos días, en línea con Hidalgo y García Mayo (2019). Por otra, IFF+Colab solo fue superior a Colab con respecto a los episodios sobre 3S. Cabe pensar, pues, que el bajo número de formas POSS en el texto de la dictoglosia (Ishikawa & Révész, 2020) y una probable menor exposición previa a la misma (Horst et al., 2010) socavaron la posibilidad de focalizar la atención a esta forma durante la tarea a pesar de haber recibido una formación específica sobre ella.

Los resultados también son alentadores en lo que se refiere a las dinámicas de pareja durante la tarea, puesto que la mayor parte de las interacciones se enmarcaron dentro de los patrones colaborativos, revelando una realidad distinta a la predominancia de patrones cooperativos observada en el estudio de Azkarai y Kopinska (2020). Cabe destacar que el hecho de que IFF+Colab manifestara una mayor proporción de dinámicas no colaborativas pudo radicar no tanto en el hecho de la instrucción en sí como en las características de su condición experimental. Estas consistieron en grabar a más de una pareja por turno y un menor escrutinio del investigador, lo que resultó en un entorno más parecido de su aula ILE habitual.

En cuanto al efecto del trabajo colaborativo en la escritura, el presente estudio no ofrece indicios de una superioridad de la escritura en parejas frente a la individual, lo que concordaría con la tesis de Elabdali (2021) sobre una menor influencia de la colaboración en tareas que implican una reproducción de un texto, como la dictoglosia. Aun así, la combinación de la RT y la colaboración permitió observar una tendencia a incrementar la corrección gramatical de las reconstrucciones el segundo día, mientras que esta dimensión permaneció inalterada en el grupo Comp. Por otra parte, la IFF pareció tener un efecto (aunque no significativo) en el índice de corrección de las formas meta (3S y POSS), frente a la de Colab y IFF+Colab.

Respecto al impacto de la tarea en la calidad de la producción escrita individual, nuestra investigación tampoco ofrece claras muestras de cambios en las distintas dimensiones de escritura en L2, lo que nos lleva a una reflexión sobre las medidas empleadas (Hidalgo & Lázaro Ibarrola, 2020) y sobre la capacidad de una breve intervención de escritura colaborativa para generar cambios positivos (Nitta & Baba, 2014). Esto último adquiere si cabe más relevancia dada la gran variabilidad intrínsecamente asociada a la población infantil (Butler, 2019).

El estudio de las actitudes de los participantes también arroja resultados claramente positivos, ya que se muestran favorables a dos componentes principales de este tipo de tareas (la escritura y el trabajo colaborativo) y, en línea con trabajos anteriores (Calzada & García Mayo, 2020b; Kopinska & Azkarai, 2020; Shak, 2006), percibieron positivamente la propia dictoglosia. Por último, la mayor contribución de las variables afectivas en los logros de escritura temprana en L2 pone en valor la importancia de desarrollar experiencias positivas hacia el aprendizaje de ILE en Educación Primaria (Mihaljević Djigunović & Nikolov, 2019).

Nuestro estudio pretende, asimismo, informar las prácticas pedagógicas de los profesores de ILE en esta etapa educativa. El balance positivo que se desprende de estos resultados puede servir de motivación para implementar la tarea de la dictoglosia en sus aulas. Cabe recordar varios aspectos cruciales que garanticen una mayor efectividad de la tarea en su objetivo de focalizar la atención de los aprendices hacia la forma, como por ejemplo, que el vocabulario del texto sea lo suficientemente conocido (y se reduzca así la carga cognitiva de la tarea), que estén familiarizados con el procedimiento (por ejemplo, llevándola a cabo con anterioridad en su L1) y que las formas gramaticales meta sean conocidas y adecuadas para su etapa de desarrollo. Aparte de eso, dada la acogida positiva de la tarea y la mayor calidad lingüística exhibida en las reconstrucciones frente a la narración en L2 más libre (como indicaron los pretests y postests), implementando de forma habitual la dictoglosia en el aula se podría promover la percepción de autoeficacia de los aprendices y generar actitudes más favorables hacia la escritura en L2 (Schunk, 1984), que, a su vez, repercutieran positivamente en la competencia escrita en este idioma.

En suma, la presente tesis ha contribuido a comprender mejor los procesos, productos y percepciones asociados a la tarea de la dictoglosia en niños de 11-12 años que aprenden ILE. Al tratarse de un ámbito de investigación incipiente, esperamos que los resultados obtenidos sirvan para establecer futuras líneas de investigación en el campo de la escritura colaborativa temprana en L2 y para orientar la labor de los profesionales de la enseñanza de lenguas en Educación Primaria.

Appendices

APPENDIX A

A1 Language background questionnaire in Spanish

LASLAB

Universidad del País Vasco — Euskal Herriko Unibertsitatea

Cuestionario sobre el conocimiento de lenguas

- **Nombre y apellidos:** ____
- Clase: _____________
- Edad: ______________
- Género: Masculino ☐ Femenino ☐

- Haz una lista de todas las lenguas que conoces
  Pon un ✔ a la(s) lengua(s) que hablas <u>en casa</u>
  1. ________________________
  2. ________________________
  3. ________________________
  4. ________________________
  5. ________________________

- **En qué lenguas puedes…**
  Leer: ____
  Hablar: ____
  Escribir: ____

- **Qué lengua(s) utilizas…**
  En la escuela: ____
  Con los amigos: ____
  Cuando sueñas: ____

- ¿Cuántos <u>años</u> has estudiado **inglés**?
  ____

- ¿Cuántas <u>horas</u> de **inglés** tienes <u>a la semana</u>? (incluyendo las horas extraescolares, como la academia o las clases particulares)
  ____

- ¿Has vivido en algún país de habla inglesa? ¿Dónde? ¿Durante cuánto tiempo?
  ____

- ¿Has aprendido o estás aprendiendo alguna otra lengua extranjera aparte del inglés? ¿Dónde? ¿Cuántas horas a la semana?
  ____

¡Gracias por tu ayuda! Eskerrik asko zure laguntzagatik!

Thank you for your help!

A2 Language background questionnaire in English

LASLAB

Universidad del País Vasco Euskal Herriko Unibertsitatea

Language background questionnaire

- **Name and surnames:** ______________________________
- Class: __________
- Age: __________
- Sex: Male ☐ Female ☐

- Make a list of **all the languages you know**
  Put a ✔ next to the language(s) you speak at home
  1. ________________
  2. ________________
  3. ________________
  4. ________________
  5. ________________

- **In what languages can you…**
  Read: ______________________________
  Speak: ______________________________
  Write: ______________________________

- **What language(s) do you speak…**
  At school: ______________________________
  With your friends: ______________________________
  When you dream: ______________________________

- How many <u>years</u> have you studied English?
  ____

- How many <u>hours</u> of **English** do you have <u>per week</u>? (including those outside school, such as classes at the academy or private lessons)
  ____

- Have you ever lived in an English-speaking country? Where? How long?
  ____

- Have you ever learnt or are you learning now a foreign language apart from English? Where? How many hours a week?
  ____

¡Gracias por tu ayuda! Eskerrik asko zure laguntzagatik!

Thank you for your help!

A3 Experimental dictogloss texts and worksheets (DG1 and DG2)

Naughty Laura (DG1)

Laura takes her lunch box everyday to school. Today, her father forgot to prepare a sandwich, so he gives her some money to buy an apple at the supermarket. At the supermarket, she sees some chocolate bars. She loves chocolate, so she buys one with black chocolate and peanuts instead of an apple. At the break she feels very hungry and eats the chocolate bar. It tastes so good! Then, she returns to class. They have Maths. Suddenly, her face turns red and she starts to feel very sick. Laura forgot she is allergic to peanuts! The teacher calls her father and he drives her to hospital. At the hospital, her father tells her: “an apple a day keeps the doctor away”.

122 words, 15 instances of 3S

A celebration (DG2)

The Smiths are spending a day together in the garden. Tom is playing football with his uncle. He calls his grandmother to join them, but she is busy playing cards and she is winning all the time! Now it is lunchtime. They are all sitting at the table in the garden. María, the oldest granddaughter, asks her mum to take some pictures. Her dad pulls funny faces and they laugh. In the afternoon, it’s very hot. Tom sees that his grandfather is preparing ice cream in the kitchen. He calls his sister and they go quickly to try it. It’s delicious! María gives a big hug to her grandfather. Then, she takes a portion for her aunt, who is in the swimming pool. What a wonderful day!

127 words, 8 instances of *POSS*

Figure 36 DG1 note-taking worksheet

LASLAB

Universidad del País Vasco · Euskal Herriko Unibertsitatea

T1

Nombre y apellidos: ______________________________ **Clase:**______________

Usa este espacio para tomar notas la segunda vez que escuches el texto. La imagen puede servirte de ayuda:

Use this space to take notes while you listen to the text a second time. The picture can help you:

Figure 37 DG1 reconstruction worksheet

LASLAB

Universidad del País Vasco · Euskal Herriko Unibertsitatea

T1

1) Nombre y apellidos: ______________________________ **Clase:**______________

2) Nombre y apellidos: ______________________________ **Clase:**______________

Ahora tenéis que escribir el texto. Recordad que debe ser <u>lo más parecido posible</u> al que acabáis de escuchar.

Now you have to write your text. Remember that it should be <u>as similar as possible</u> to the one you have just listened to.

THANK YOU FOR YOUR TIME! 😊

Figure 38 DG2 note-taking worksheet

LASLAB   Universidad del País Vasco / Euskal Herriko Unibertsitatea

T1

Nombre y apellidos: ______________________________ **Clase:**______________

Usa este espacio para tomar notas la segunda vez que escuches el texto. La imagen puede servirte de ayuda:

Use this space to take notes while you listen to the text a second time. The picture can help you:

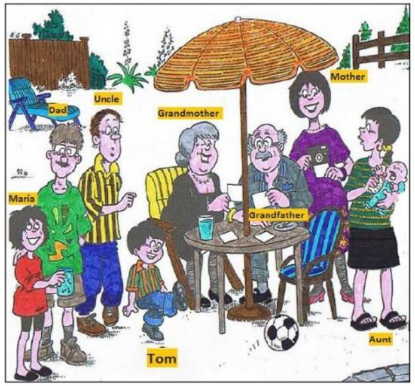


Figure 39 DG2 reconstruction worksheet

LASLAB   Universidad del País Vasco / Euskal Herriko Unibertsitatea

T1

1) Nombre y apellidos: ______________________________ **Clase:**______________

2) Nombre y apellidos: ______________________________ **Clase:**______________

Ahora tenéis que escribir el texto. Recordad que debe ser <u>lo más parecido posible</u> al que acabáis de escuchar.

Now you have to write your text. Remember that it should be <u>as similar as possible</u> to the one you have just listened to.

THANK YOU FOR YOUR TIME! 😊

A4 Pilot dictogloss text materials

Diktoglosia: Amona Joxepa

Amona Joxepa, amona guztiak bezala, heldua da baina ez da zaharra. Ile zuria du, eguzki izpiak bezain distiratsuak, eta beti oso ongi orraztua eramaten du.

Gauero, oheratzerakoan, istorio zoragarriak kontatzen dizkit, eta bere ahots atsegin eta goxoa entzunez lokartzen naiz. Goizero, berak egiten dizkidan pastel goxoen usainaz esnatzen naiz.

Baina nik aurten bere urtebetetzerako sorpresatxo bat prestatuko diot. Bera baino lehenago jaikiko naiz eta gosaria prestatuko diot: kafesnea, bere gozotegi gogokoenean erositako kruasana eta zukua. Bere logelara joango naiz eta zorionak zuri abestuko diot. Ziur oso pozik jarriko dela eta besarkada handi bat emango didala, ez uste?

Figure 40 Pilot dictogloss note-taking worksheet

LASLAB

Universidad del País Vasco Euskal Herriko Unibertsitatea

T0

Izen abizenak: ________________________________ **Taldea:**______________

Erabili kuadro hau istorioaren hitz gakoak apuntatzeko:

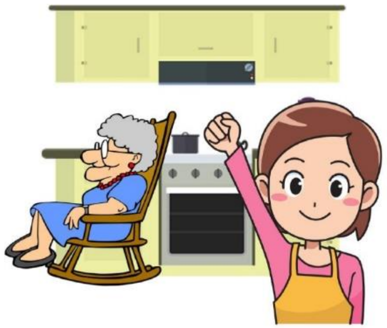


Figure 41 Pilot dictogloss reconstruction worksheet

LASLAB

Universidad del País Vasco Euskal Herriko Unibertsitatea

T0

Izen abizenak: ________________________________ **Taldea:**______________

Orain testua berridatzi behar duzu. Gogoratu entzun duzunaren <u>ahalik eta antzekoena</u> izan behar duela.

Figure 42 Written languaging worksheet pilot dictogloss

LASLAB

Universidad del País Vasco Euskal Herriko Unibertsitatea

T0

Izen abizenak: ______________________________ Taldea:______________

Orain erabili ezazu kuadro hau idazterakoan izan dituzun zalantzak eta galderak jartzeko. Hurrengo esaldiak eredu bezala erabil ditzakezu edo, nahiago izanez gero, zure hitzak erabili.

- **Ez dakit ………………….. nola esaten/idazten den euskaraz.**
- **……………….. idatzi dut baina ez nago ziur zuzena ote den.**
- **Nola esaten da ……………………. aditza etorkizunean?**
- **Ez dakit testuak …………………………….. edo ……………………. esaten zuen.**
- **Ziur nago ……………………… horrela idazten dela.**

Eskerrik asko zure laguntzagatik!

A5 FFI instruction materials

Figure 43 Pre-DG1 FFI materials on 3S

LASLAB

Universidad del País Vasco — Euskal Herriko Unibertsitatea

Nombre y apellidos: ______________________________ **Clase:** ______________

¡Un montón de serpientes andan sueltas por el texto! Lee la historia y piensa de qué regla gramatical se trata. Coméntalo con tu compañero/a y, por último, pondremos todo en común.

A lot of snakes are on the loose in the text! Read the text and think about what grammar rule it can be. Discuss it with your partner and, finally, we will put everything in common.

A sweet surprise

Next Sunday **Mary's grandmother** **celebrateS** her birthday. **Her grandmother** always **cookS** delicious things for her but once a year **Mary likeS** giving her a sweet surprise. **She wakeS** up early in the morning and **buyS** the ingredients at the supermarket. At home, first, **she putS** sugar and some flour in a bowl. Then, **she breakS** some eggs and **beatS** them. **She** also **addS** some milk. Her brother **Tom helpS** her to put the mixture in muffin cups and they bake them in the oven. Finally, **he pourS** melted chocolate and sweets on top of the cupcakes, because **her granny loveS** them.

At 6 o'clock, Mary and Tom visit their granny. After trying a cupcake, **her granny sayS**: "**It tasteS** delicious!"

Ahora tienes que completar la regla de oro.

Now you have to complete the golden rule.

When does the S appear?

Hint: when the subject is

______, ______ or ______

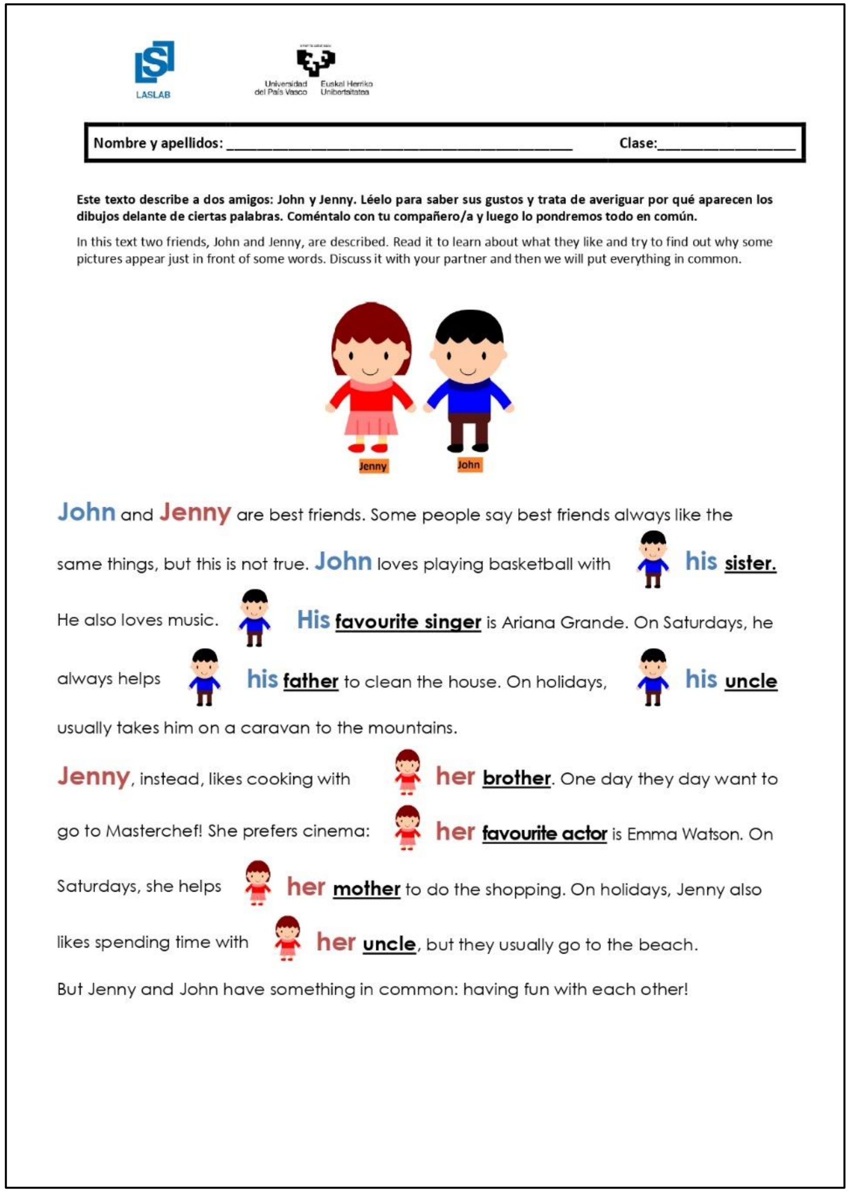
Por último, pon las serpientes en este texto cuando creas que es necesario. Subraya los sujetos como en el texto modelo.

Finally, put the snakes in this text when you think it's necessary. Underline the subjects following the sample text.

Jing's life

Jing is 16 years old and she is a kung fu student in China. Kung fu students usually take lessons in the morning and kung fu training in the afternoon and evening. There are about 4.000 students in her school! Her day start at 5:00 in the morning and end at 9:30. She sleep in the school. She only see her parents twice a year, because they live very far from the school. On Sundays, her free day, she talk to her friends, play video games and listen to music. She has got a lot of friends in the school, they are a big family. She love her school because they do kung fu everyday. In the future, she want to travel around the world, and visit the important cities in Europe. Her dream is to visit the Eiffel tower in Paris. She also want to make kung fu films one day!

Figure 44 Pre-DG2 FFI materials on POSS



LASLAB

Universidad del País Vasco Euskal Herriko Unibertsitatea

Nombre y apellidos: ________________________ **Clase:**____________

Este texto describe a dos amigos: John y Jenny. Léelo para saber sus gustos y trata de averiguar por qué aparecen los dibujos delante de ciertas palabras. Coméntalo con tu compañero/a y luego lo pondremos todo en común.

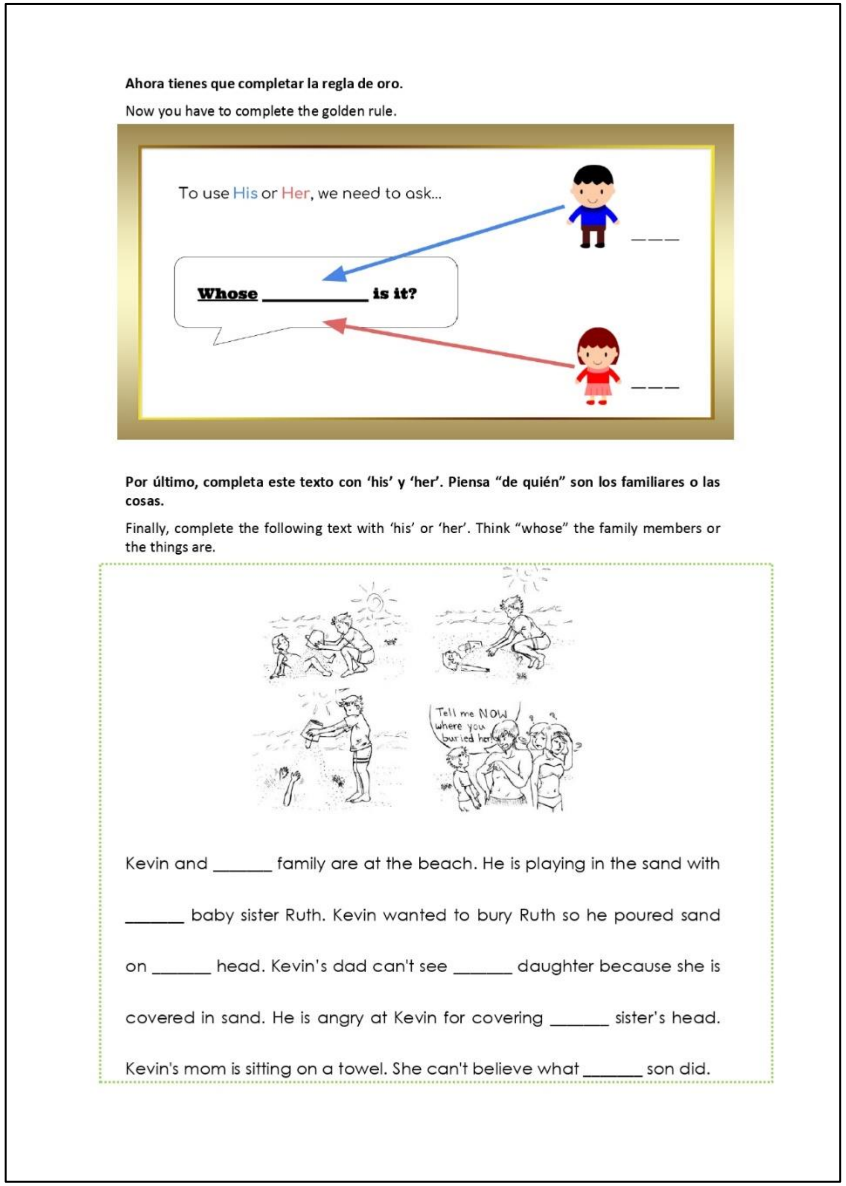
In this text two friends, John and Jenny, are described. Read it to learn about what they like and try to find out why some pictures appear just in front of some words. Discuss it with your partner and then we will put everything in common.

Jenny John

John and Jenny are best friends. Some people say best friends always like the same things, but this is not true. John loves playing basketball with his **sister.** He also loves music. His **favourite singer** is Ariana Grande. On Saturdays, he always helps his **father** to clean the house. On holidays, his **uncle** usually takes him on a caravan to the mountains.

Jenny, instead, likes cooking with her **brother**. One day they day want to go to Masterchef! She prefers cinema: her **favourite actor** is Emma Watson. On Saturdays, she helps her **mother** to do the shopping. On holidays, Jenny also likes spending time with her **uncle**, but they usually go to the beach.

But Jenny and John have something in common: having fun with each other!



Ahora tienes que completar la regla de oro.

Now you have to complete the golden rule.

To use His or Her, we need to ask...

Whose __________ **is it?**

___

___

Por último, completa este texto con 'his' y 'her'. Piensa "de quién" son los familiares o las cosas.

Finally, complete the following text with 'his' or 'her'. Think "whose" the family members or the things are.

Tell me NOW where you buried her

Kevin and ______ family are at the beach. He is playing in the sand with ______ baby sister Ruth. Kevin wanted to bury Ruth so he poured sand on ______ head. Kevin's dad can't see ______ daughter because she is covered in sand. He is angry at Kevin for covering ______ sister's head. Kevin's mom is sitting on a towel. She can't believe what ______ son did.

A6 Individual languaging worksheet (Comp)

LASLAB

Universidad del País Vasco Euskal Herriko Unibertsitatea

T1

Nombre y apellidos: ____ **Clase:**_________________

Ahora escribe en esta página qué <u>problemas o dudas</u> has tenido al escribir el texto. Puedes utilizar las siguientes frases como modelo, o si lo prefieres, puedes usar tus propias palabras.

- **No sé cómo se dice/escribe ……………………………….. en inglés**
- **He escrito ……………………………………….. así pero no estoy seguro/a de que esté bien.**
- **¿Cómo se dice …………………………….. en pasado?**
- **No sé si en el texto se decía …………………………………….. o …………………………………………..**
- **Estoy seguro/a de que …………………………………….. se escribe así**

¡MUCHAS GRACIAS POR TU TIEMPO!

THANK YOU FOR YOUR TIME!

A7 L2 writing tests

Figure 45 L2 writing test– Version 1

Name and surname: ____________________ **Class:** ______ **T1**

Writing in English 1

You are going to write a story about a holiday. You are given the beginning and you have to invent the rest of it. Mind your handwriting!

The Smiths are at New York airport waiting for their flight to London. Tom and Kevin are travelling with their mother. They are going to see their uncle. **Kevin** is listening to music but suddenly… ____________

The end

1

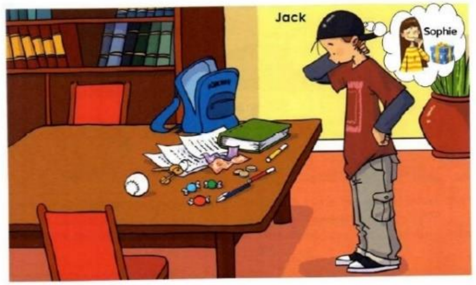
Now turn the page

Figure 46 L2 writing test – Version 2

Name and surname: ____________________ **Class:** ______ **T1**

Writing in English 2

You are going to write a story about a boy who loses something important. You are given the beginning and you have to invent the rest of it. Mind your handwriting!



Jack and Sophie have been friends for a long time. Today **Jack** is very excited because Sophie celebrates her birthday. He is searching in his backpack but something is wrong… ____________

The end

1

Now turn the page

Figure 47 L2 writing test– Version 3

Name and surname: ______________________ **Class:**________ | **T1**

Writing in English 3

You are going to write a story about an amazing discovery. You are given the beginning and you have to invent the rest of it. Mind your handwriting!

Today Sarah has invited **Katie** to come with their family to the forest. While she is preparing lunch, Katie sees something in the lake… ______________

The end

1

Now turn the page

A8 L1 writing tests

Figure 48 L1 writing test - Version 1

Nombre y apellidos: ____________________ Clase:______ | T1

Escribir en castellano 1

Vas a escribir una historia sobre una avería que trastoca los planes de Xabier. Tienes el comienzo y ahora te toca a ti inventar el resto.

Procura escribir con letra clara y fácil de leer

Xabier y su equipo tienen que viajar a Madrid para jugar la gran final del torneo de fútbol de primavera. Sin embargo, el día que tienen que irse en la furgoneta de los padres de Xabier algo no va como debería...

1

Tienes más espacio por la otra cara

Figure 49 L1 writing test – Version 2

Nombre y apellidos: ____________________ Clase:______ | T1

Escribir en castellano 2

Vas a escribir una historia sobre un regalo inesperado. Tienes el comienzo y ahora te toca a ti inventar el resto.

Procura escribir con letra clara y fácil de leer

Hoy es el cumpleaños de Miren. Cuando se despierta encuentra una vieja alfombra y una nota escrita a mano que dice "¡*Zorionak*, Miren! Ahora te toca a ti vivir una gran aventura, te presto mi alfombra voladora. ¡Disfrútala!"

1

Tienes más espacio por la otra cara

Figure 50 L1 writing test – Version 3

Nombre y apellidos: ______________________________ Clase:______	T1

Escribir en castellano 3

Vas a escribir una historia de misterio. Tienes el comienzo y ahora te toca a ti inventar el resto.

Procura escribir con letra clara y fácil de leer ✍

Ane se ha quedado sola en casa esta tarde. Está tranquilamente viendo la televisión en el salón cuando de pronto escucha un ruido que viene de la habitación____

1

Tienes más espacio por la otra cara

A9 AQ - Pretask attitude questionnaire (Spanish version)

Nombre y Apellidos: ______________________ **Clase:** ______

Cuestionario de actitudes (1)

Primera parte

Marca las caras dependiendo de si estás de acuerdo o no con las afirmaciones.

¡Claro que NO! | No | No estoy seguro/a | Sí | ¡Claro que SÍ!

*Por ejemplo, si te gustan **mucho** los macarrones deberás marcarlo así*

Me gustan los macarrones

*Si te gustan **bastante**, así*

Me gustan los macarrones

*Y si no te gustan **nada**, así*

Me gustan los macarrones

¡Recuerda! No hay respuestas correctas ni incorrectas, solo nos interesa saber tu opinión

A. Escribir en castellano

1. Me siento cómodo/a al escribir en castellano
2. Me pone nervioso/a tener un límite de tiempo para escribir en castellano
3. Si mi redacción en castellano va a ser para nota, me pongo nervioso/a
4. Me preocupa lo que mis compañeros/as piensen de mis redacciones en castellano
5. Siempre me parece que escribo peor en castellano que mis compañeros/as
6. Cuando escribo en castellano, puedo expresar mis ideas con facilidad sin salirme del tema
7. Una vez terminada la redacción en castellano, siempre la repaso
8. Soy capaz de encontrar errores en mi redacción en castellano y de corregirlos
9. Dedico tiempo suficiente a escribir las redacciones en castellano que me mandan hacer en clase
10. Soy capaz de escribir un texto correctamente organizado en castellano

1

Sigue en la página 2

¡Claro que NO! | No | No estoy seguro/a | Sí | ¡Claro que SÍ!

B. Escribir en inglés

11. Escribir en inglés me pone nervioso/a
12. Estoy tranquilo/a cuando tengo un tiempo límite para escribir en inglés
13. Estoy tranquilo/a cuando lo que escribo en inglés me lo van a corregir para evaluarme
14. Me da igual lo que mis compañeros/as piensen de mis redacciones en inglés
15. Siempre me parece que escribo peor en inglés que mis compañeros/as
16. Al escribir en inglés, se me ocurren ideas sobre el tema que toca
17. Siempre repaso mis redacciones en inglés después de terminarlas
18. Me cuesta ver los fallos en mis redacciones en inglés
19. Intento hacer las redacciones en inglés que tenemos que hacer en clase lo más rápido posible
20. Me cuesta escribir un texto en inglés con las ideas organizadas

C. Trabajo colaborativo y trabajo individual

ANTES DE RESPONDER *Nunca **(1)**, Pocas veces **(2)**, Algunas veces **(3)**, Casi siempre **(4)**, Siempre **(5)***

*¿Cada cuánto sueles trabajar **individualmente** en clase?*	1	2	3	4	5
*¿Y en **parejas**?*	1	2	3	4	5

21. Trabajar en parejas me parece muy útil
22. Me gusta trabajar con otros compañeros/as
23. En tareas difíciles prefiero trabajar solo/a
24. Aprendo más trabajando con otros compañeros/as que solo/a
25. Me concentro más trabajando solo/a que con otros compañeros/as
26. Prefiero elegir yo con quién voy a trabajar en pareja
27. Me siento a gusto trabajando en pareja con compañeros que no he elegido yo
28. Me gustaría trabajar más en parejas en clase

2

Sigue en la página 3

¡Claro que NO! | No | No estoy seguro/a | Sí | ¡Claro que SÍ!

D. Escritura colaborativa y escritura individual

ANTES DE RESPONDER *Nunca **(1)**, Pocas veces **(2)**, Algunas veces **(3)**, Casi siempre **(4)**, Siempre **(5)***

*¿Cada cuánto sueles escribir **individualmente** en clase?*	1	2	3	4	5
*¿Y en **parejas**?*	1	2	3	4	5

29. Escribiendo individualmente se me ocurren más ideas que en parejas

30. Las redacciones que hacemos en parejas son más correctas que las que hago yo solo/a

31. Me pone nervioso/a compartir mis ideas cuando escribimos en parejas

32. Estoy más tranquilo cuando tengo que escribir un texto individualmente que con algún/a compañero/a

33. Suelo prestar más atención al texto cuando escribo individualmente que en parejas

34. Cuando escribimos algo en parejas, suelo hablar de otras cosas que no son el texto

35. Aprendo escribiendo con otros compañeros/as

36. Creo que puede salir una redacción mejor escribiendo en pareja que escribiendo yo solo/a

E. Aprender inglés

37. Me gustan mis clases de inglés en la escuela

38. Me gusta usar inglés en clase

39. Mis padres creen que debería dedicar más tiempo al inglés

40. Mis padres me animan a que aprenda inglés

41. Siempre me parece que otros/as alumnos/as saben más inglés que yo

42. Me siento más tenso/a y nervioso/a en mis clases de inglés que en las de castellano

43. El inglés me ayudará a hablar con gente de otros países y otras culturas

44. Estoy aprendiendo inglés para entender vídeos, música, juegos y chats en internet

3

Sigue en la página 4

Parte 2. Para terminar…

Escribe tu respuesta con letra clara y legible

- Escribe una palabra para describir qué sientes cuando tienes que **escribir en castellano:**

____

¿Por qué has elegido esa palabra?

- Escribe una palabra para describir qué sientes cuando tienes que **escribir en inglés:**

____

¿Por qué has elegido esa palabra?

- Escribe una palabra para describir qué sientes cuando tienes que **trabajar en parejas:**

____

¿Por qué has elegido esa palabra?

- Escribe una palabra para describir qué sientes cuando tienes que **trabajar individualmente:**

____

¿Por qué has elegido esa palabra?

¡MUCHAS GRACIAS POR TU TIEMPO!

THANK YOU FOR YOUR TIME!

A10 AQ - Pretask attitude questionnaire (English version)

Name and surnames: ______________________________ **Class:** ________

Attitude Questionnaire (1)

First part

Choose one face depending on how much you agree with the following statements.

Definitely NO! | **No** | **Not sure** | **Yes** | **Definitely YES!**

*For example, if you like pasta **very much**, you should choose the following face*

I like pasta

*If you **quite** like it, this face*

I like pasta

*If you **don't like it at all**, this other face*

I like pasta

Remember! There aren't right or wrong answers, we are just interested in your opinion

A. Writing in Spanish
1. I feel comfortable writing in Spanish
2. Having a time limit when writing in Spanish makes me nervous
3. If my composition in Spanish is going to be graded, I get nervous
4. I feel concerned about what my classmates may think about my compositions in Spanish
5. I always think that I write worse in Spanish than my classmates
6. When I write in Spanish, I can easily convey my ideas without going off the point
7. Every time I finish a composition in Spanish, I always revise it
8. I am capable of finding and correcting mistakes in my Spanish composition
9. I devote enough time to the Spanish compositions they make me write in class
10. I can write a correctly organized text in Spanish

1

Continue to page 2

Definitely NO! | **No** | **Not sure** | **Yes** | **Definitely YES!**

B. Writing in English
11. Writing in English makes me nervous
12. I feel calm when I have a time limit to write in English
13. I feel calm when my English composition is going to be graded
14. I don't care what my classmates think about my English compositions
15. I always think that I write worse in English than my classmates
16. When writing in English, I can always come up with ideas about the composition topic
17. After finishing my compositions in English, I always revise them
18. I have trouble in finding mistakes in my English compositions
19. I always try to do the English compositions we have to write in class as fast as possible
20. I find it hard to write a composition in English with organized ideas

C. Collaborative work and individual work

BEFORE ANSWERING *Never **(1)**, Hardly ever **(2)**, Sometimes **(3)**, Very often **(4)**, Always **(5)***

How often do you work individually in class?	1	2	3	4	5
And in pairs?	1	2	3	4	5

C. Collaborative work and individual work
21. I think working in pairs is useful
22. I like working with my peers
23. I prefer working individually in difficult tasks
24. I learn more working with my peers than on my own
25. I focus more when I am working on my own than with other peers
26. I prefer to be the one choosing who I get to work with in pairs
27. I feel comfortable working in pairs with peers than I have not chosen
28. I'd like to work more in pairs in class

2

Continue to page 3

¡Claro que NO!	No	No estoy seguro/a	Sí	¡Claro que SÍ!

D. Collaborative writing and individual writing

BEFORE ANSWERING *Never* ***(1)****, Hardly ever* ***(2)****, Sometimes* ***(3)****, Very often* ***(4)****, Always* ***(5)***

*How often do you write **individually** in class?*	1	2	3	4	5
And in pairs?	1	2	3	4	5

29. I can come up with more ideas when writing individually than in pairs

30. The compositions that we write in pairs are more accurate than the ones I write individually

31. I feel nervous when I have to share my ideas when writing in pairs

32. I feel calmer when I have to write a text individually than in pairs

33. I usually pay more attention to the text when I am writing individually than in pairs

34. When we write something in pairs, I speak about things that are not related to the text

35. I learn things when writing with other peers

36. I think that the text can be better when writing in pairs than on my own

E. Learning English

37. I like my English lessons at school

38. I like using English in class

39. My parents think that I should devote more time to English

40. My parents encourage me to learn English

41. I always think that other students know more English than I do

42. I feel more tense and nervous in my English class than in the Spanish one

43. English will help me speak with people from other countries and cultures

44. I am learning English to understand videos, music, games and chats on the Internet

Continue to page 4

Part 2. To finish…

Write your answer with a neat and clear handwriting

- Write one word to describe how you feel when you **write in Spanish:**

____

Why did you choose that word?

- Write one word to describe how you feel when you **write in English:**

____

Why did you choose that word?

- Write one word to describe how you feel when you have to **work in pairs:**

____

Why did you choose that word?

- Write one word to describe how you feel when you **work individually:**

____

Why did you choose that word?

¡MUCHAS GRACIAS POR TU TIEMPO!

THANK YOU FOR YOUR TIME!

A11 DQ - Posttask attitude questionnaire – Comp and Collab (Spanish version)

¡Claro que NO! | No | No estoy seguro/a | Sí | ¡Claro que SÍ!

SOLO SI HAS REALIZADO LAS TAREAS EN PAREJA

C. La tarea en parejas

15. Me gustaría volver a hacer las tareas en mi clase habitual con mi profesor/a tal y como las he hecho ahora
16. Si hacemos las tareas otra vez, quiero seguir estando en parejas
17. He estado a gusto trabajando con mi compañero/a
18. Las tareas me han parecido demasiado difíciles
19. La parte de escuchar el audio me ha gustado
20. La parte de tomar apuntes mientras escucho el audio me ha gustado
21. La parte de reescribir el texto me ha gustado
22. La parte de hablar con mi compañero/a me ha gustado
23. Creo que se me ha dado mejor la tarea el segundo día que el primero
24. Cuantas más veces haga la tarea, mejor se me dará
25. Me he sentido más motivado/a para hacer la tarea el segundo día que el primero

3

Sigue en la página 4 si hiciste la tarea INDIVIDUALMENTE

¡Claro que NO! | No | No estoy seguro/a | Sí | ¡Claro que SÍ!

SOLO SI HAS REALIZADO LAS TAREAS INDIVIDUALEMENTE

D. La tarea individual

15. Me gustaría volver a hacer las tareas en mi clase habitual con mi profesor/a tal y como las he hecho ahora
16. Volvería a hacer las tareas individualmente
17. Me he sentido cómodo trabajando solo/a
18. Las tareas me han parecido demasiado difíciles
19. La parte de escuchar el audio me ha gustado
20. La parte de tomar apuntes mientras escucho el audio me ha gustado
21. La parte de reescribir el texto me ha gustado
22. La parte de anotar mis dudas y reflexiones en un papel me ha gustado
23. Creo que se me ha dado mejor la tarea el segundo día que el primero
24. Cuantas más veces haga la tarea, mejor se me dará
25. Me he sentido más motivado/a para hacer la tarea el segundo día que el primero

4

Sigue en la página 5

Para terminar…

Escribe tus respuestas con letra clara y legible

Fíjate bien en lo que te pide cada pregunta:

1. Escribe una cosa que te haya gustado de **las tareas:**

____

¿Por qué has elegido esa palabra?

____

____

2. Escribe una cosa que te **NO** haya gustado de **las tareas:**

____

¿Por qué has elegido esa palabra?

____

____

SOLO SI HAS REALIZADO LAS TAREAS EN PAREJA

3. Escribe una cosa que te haya gustado de **trabajar en parejas:**

____

¿Por qué has elegido esa palabra?

____

____

4. Escribe una cosa que **NO** te haya gustado de **trabajar en parejas:**

____

¿Por qué has elegido esa palabra?

____

____

Sigue en la página 6

SOLO SI HAS REALIZADO LAS TAREAS INDIVIDUALEMENTE

3. Escribe una cosa que te haya gustado de **trabajar individualmente:**

____

¿Por qué has elegido esa palabra?

____

____

4. Escribe una cosa que **NO** te haya gustado de **trabajar individualmente:**

____

¿Por qué has elegido esa palabra?

____

____

5. Escribe una palabra para describir qué sientes cuando tienes que escribir en **castellano:**

____

¿Por qué has elegido esa palabra?

____

____

6. Escribe una palabra para describir qué sientes cuando tienes que escribir en **inglés:**

____

¿Por qué has elegido esa palabra?

____

____

¡MUCHAS GRACIAS POR TU TIEMPO! / ESKERRIK ASKO ZURE DENBORAGATIK!

THANK YOU FOR YOUR TIME!

A12 DQ - Posttask attitude questionnaire – Comp and Collab (English version)

Definitely NO!	No	Not sure	Yes	Definitely YES!

ONLY IF YOU DID THE TASKS IN PAIRS

C. The task in pairs

15. I'd like to do the tasks again in the normal lessons with my teacher just as I did them now
16. If we do the tasks again, I'd like to be in pairs again
17. I felt at ease working with my peer
18. I found the tasks too difficult
19. I enjoyed the listening part
20. I enjoyed the part where I had to take down notes while I listened to the audio
21. I enjoyed the rewriting part
22. I enjoyed the part where I had to speak to my peer
23. I think that I did the task better on the second day than on the first
24. The more I do the task, the better I'll be at it
25. I felt more motivated to do the task on the second day than on the first

✎ *Continue to page 4 if you did the task INDIVIDUALLY*

Definitely NO!	No	Not sure	Yes	Definitely YES!

ONLY IF YOU DID THE TASKS INDIVIDUALLY

D. The individual task

15. I'd like to do the tasks again in the normal lessons with my teacher just as I did them now
16. I'd do the tasks individually again
17. I felt at ease working on my own
18. I found the tasks too difficult
19. I enjoyed the listening part
20. I enjoyed the part where I had to take down notes while I listened to the audio
21. I enjoyed the rewriting part
22. I enjoyed the part where I had to write down my doubts and thoughts on a photocopy
23. I think that I did the task better on the second day than on the first
24. The more I do the task, the better I'll be at it
25. I felt more motivated to do the task on the second day than on the first

✎ *Continue to page 5*

To finish…

Write your answer with a neat and clear handwriting ✍

Pay attention to what each question asks:

1. Write one thing you enjoyed about **the tasks:**

Why did you choose that word?

2. Write one thing you **DID NOT** enjoy about **the tasks:**

Why did you choose that word?

ONLY IF YOU DID THE TASKS IN PAIRS

3. Write one thing you enjoyed about **working in pairs:**

Why did you choose that word?

4. Write one thing you **DID NOT** enjoy about **working in pairs:**

Why did you choose that word?

5

Continue to page 6

ONLY IF YOU DID THE TASKS INDIVIDUALLY

3. Write one thing you enjoyed about **working individually:**

Why did you choose that word?

4. Write one thing you **DID NOT** enjoy about **working individually:**

Why did you choose that word?

5. Write one word to describe how you feel when you write in **Spanish:**

Why did you choose that word?

6. Write one word to describe how you feel when you write in **English:**

Why did you choose that word?

¡MUCHAS GRACIAS POR TU TIEMPO! / ESKERRIK ASKO ZURE DENBORAGATIK!

THANK YOU FOR YOUR TIME!

6

A13 DQ - Posttask attitude questionnaire – FFI+Collab (Spanish version)

¡Claro que NO! | No | No estoy seguro/a | Sí | ¡Claro que SÍ!

C. La tarea en parejas					
15. Me gustaría volver a hacer las tareas en mi clase habitual con mi profesor/a tal y como las he hecho ahora					
16. Si hacemos las tareas otra vez, quiero seguir estando en parejas					
17. He estado a gusto trabajando con mi compañero/a					
18. Las tareas me han parecido demasiado difíciles					
19. Los ejercicios de gramática antes de la tarea de escritura me han gustado					
20. La parte de escuchar el audio me ha gustado					
21. La parte de tomar apuntes mientras escucho el audio me ha gustado					
22. La parte de reescribir el texto me ha gustado					
23. La parte de hablar con mi compañero/a me ha gustado					
24. Creo que se me ha dado mejor la tarea el segundo día que el primero					
25. Cuantas más veces haga la tarea, mejor se me dará					
26. Me he sentido más motivado/a para hacer la tarea el segundo día que el primero					

3

Sigue en la página 4

Para terminar…

Escribe tus respuestas con letra clara y legible

Fíjate bien en lo que te pide cada pregunta:

1. Escribe una cosa que te haya gustado de **los ejercicios de gramática:**

____

¿Por qué has elegido esa palabra?

____

____

2. Escribe una cosa que te **NO** haya gustado de **los ejercicios de gramática:**

____

¿Por qué has elegido esa palabra?

____

____

3. Escribe una cosa que te haya gustado de **las tareas de escritura colaborativa:**

____

¿Por qué has elegido esa palabra?

____

____

4. Escribe una cosa que te **NO** haya gustado de **las tareas de escritura colaborativa:**

____

¿Por qué has elegido esa palabra?

____

____

4

Sigue en la página 5

5. Escribe una cosa que te haya gustado de **trabajar en parejas:**

____

¿Por qué has elegido esa palabra?

____

____

6. Escribe una cosa que **NO** te haya gustado de **trabajar en parejas:**

____

¿Por qué has elegido esa palabra?

____

____

7. Escribe una palabra para describir qué sientes cuando tienes que escribir en **castellano:**

____

¿Por qué has elegido esa palabra?

____

____

8. Escribe una palabra para describir qué sientes cuando tienes que escribir en **inglés:**

____

¿Por qué has elegido esa palabra?

____

____

¡MUCHAS GRACIAS POR TU TIEMPO! / ESKERRIK ASKO ZURE DENBORAGATIK!

THANK YOU FOR YOUR TIME!

A14 DQ - Posttask attitude questionnaire – FFI+Collab (English version)

C. The task in pairs

15. I'd like to do the tasks again in the normal lessons with my teacher just as I did them now
16. If we do the tasks again, I'd like to be in pairs again
17. I felt at ease working with my peer
18. I found the tasks too difficult
19. I enjoyed the grammar exercises before the writing task
20. I enjoyed the listening part
21. I enjoyed the part where I had to take down notes while I listened to the audio
22. I enjoyed the rewriting part
23. I enjoyed the part where I had to speak to my peer
24. I think that I did the task better on the second day than on the first
25. The more I do the task, the better I'll be at it
26. I felt more motivated to do the task on the second day than on the first

Continue to page 4

To finish…

Write your answer with a neat and clear handwriting

Pay attention to what each question asks:

1. Write one thing you enjoyed about **the grammar exercises:**

Why did you choose that word?

2. Write one thing you **DID NOT** enjoy about **the grammar exercises:**

Why did you choose that word?

3. Write one thing you enjoyed about **the collaborative writing tasks:**

Why did you choose that word?

4. Write one thing you **DID NOT** enjoy about **the collaborative writing tasks:**

Why did you choose that word?

Continue to page 6

5. Write one thing you enjoyed about **working in pairs:**

____

Why did you choose that word?

____

____

6. Write one thing you **DID NOT** enjoy about **working in pairs:**

____

Why did you choose that word?

____

____

7. Write one word to describe how you feel when you write in **Spanish:**

____

Why did you choose that word?

____

____

8. Write one word to describe how you feel when you write in **English:**

____

Why did you choose that word?

____

____

¡MUCHAS GRACIAS POR TU TIEMPO! / ESKERRIK ASKO ZURE DENBORAGATIK!

THANK YOU FOR YOUR TIME!

A15 Approval from Ethical Committee (UPV/EHU)

Universidad del País Vasco / Euskal Herriko Unibertsitatea

NAZIOARTEKO BIKAINTASUN CAMPUSA
CAMPUS DE EXCELENCIA INTERNACIONAL

IKERKETA SAILEKO ERREKTOREORDETZA
VICERRECTORADO DE INVESTIGACIÓN

GIZAKIEKIN ETA HAUEN LAGIN ETA DATUEKIN EGINDAKO IKERKETEI BURUZKO ETIKA BATZORDEAREN (GIEB-UPV/EHU) TXOSTENA

Mª Jesús Marcos Muñoz andreak, Universidad del País Vasco/Euskal Herriko Unibertsitateko (UPV/EHU) GIEBeko idazkari gisa,

ZIURTATZEN DU

Ezen gizakiekin egindako ikerkuntzaren etika batzorde honek, GIEB-UPV/EHU, (2014/2/17ko 32. EHAA)
Balioetsi duela ondoko ikertzailearen proposamen hau:
Asier Calzada Lizarraga andreak, M10_2018_096, honako ikerketa proiektu hau egiteko:
"La dictoglosia colaborativa en el aula de inglés como lengua extranjera en Educacion Primaria y sus efectos en la adquisión de la gramática del inglés"

Eta aintzat hartuta ezen

1. Ikerketa justifikatuta dago, bere helburuei esker jakintza areagotu eta gizarteari onura ekarriko baitio, ikerlanak lekartzakeen eragozpen eta arriskuak arrazoizko izanik.
2. Ikertzaile taldearen gaitasuna eta erabilgarri dituzten baliabideak aproposak dira proiektua gauzatzeko.
3. Ikerketaren planteamendua bat dator era honetako ikerkuntza egin ahal izateko baldintza metodologiko eta etikoekin, ikerkuntza zientifikoaren praktika egokien irizpideei jarraiki.
4. Indarreko arauak betetzen ditu, ikerketa egin ahal izateko baimenak, akordioak edo hitzarmenak barne.

Aldeko Txostena eman du 2018ko apr egin duen bileran (100/2018akta) aipatutako ikerketa proiektua ondoko ikertzaileek osatutako taldeak egin dezan:

Asier Calzada Lizarraga
María Pilar García Mayo

Eta halaxe sinatu du Leioan, 2018ko maiatzaren 17an

INFORME DEL COMITÉ DE ÉTICA PARA LAS INVESTIGACIONES CON SERES HUMANOS, SUS MUESTRAS Y SUS DATOS (CEISH-UPV/EHU)

Mª Jesús Marcos Muñoz como Secretaria del CEISH de la Universidad del País Vasco/Euskal Herriko Unibertsitatea (UPV/EHU)

CERTIFICA

Que este Comité de Ética para la Investigación con Seres Humanos, CEISH-UPV/EHU, BOPV 32, 17/2/2014,
Ha evaluado la propuesta del investigador:
D. Asier Calzada Lizarraga, M10_2018_096, para la realización del proyecto de investigación: *"La dictoglosia colaborativa en el aula de inglés como lengua extranjera en Educacion Primaria y sus efectos en la adquisión de la gramática del inglés"*

Y considerando que,

1. La investigación está justificada porque sus objetivos permitirán generar un aumento del conocimiento y un beneficio para la sociedad que hace asumibles las molestias y riesgos previsibles.
2. La capacidad del equipo investigador y los recursos disponibles son los adecuados para realizarla.
3. Se plantea según los requisitos metodológicos y éticos necesarios para su ejecución, según los criterios de buenas prácticas de la investigación científica.
4. Se cumple la normativa vigente, incluidas las autorizaciones, acuerdos o convenios necesarios para llevarla a cabo.

Ha emitido en la reunión celebrada el 26 de abril de 2018 (acta 100/2018), **INFORME FAVORABLE** a que dicho proyecto de investigación sea realizado, por el equipo investigador:

Asier Calzada Lizarraga
María Pilar García Mayo

Lo que firmo en Leioa, a 17 de mayo de 2018

Mª Jesús Marcos Muñoz
GIEB-UPV/EHUko idazkari teknikoa
Secretaria Técnica del CEISH-UPV/EHU

IIEB CEID
UNIVERSIDAD DEL PAÍS VASCO EUSKAL HERRIKO UNIBERTSITATEA

susi.marcos@ehu.es
www.ehu.es/CEID

BIZKAIKO CAMPUSA
CAMPUS DE BIZKAIA
Sarriena Auzoa, z/g
48940 LEIOA

A16 Informed consent for children's parents or legal caretakers

Universidad del País Vasco / Euskal Herriko Unibertsitatea

Departamento de Filología Inglesa y Alemana y de Traducción e Interpretación

Language and Speech laboratory

HOJA DE INFORMACIÓN Y CONSENTIMIENTO INFORMADO

(versión en castellano)

El presente formulario tiene como objeto proporcionarle la información necesaria para que decida libre y voluntariamente la participación de su hijo/a en este proyecto de tesis doctoral. Es necesario que lea detenidamente la siguiente información y que pregunte si tiene alguna duda al respecto.

CONTACTO:

- Investigador: **Asier Calzada Lizarraga /** Investigadora principal: **María del Pilar García Mayo**
- Dirección: Centro de Investigación Micaela Portilla Ikergunea 3.6. c/ Justo Vélez de Elorriaga, 01006 Vitoria-Gasteiz (Álava/Araba)
- Centro: Universidad del País Vasco / Euskal Herriko Unibertsitatea
- Teléfono: 945 01 34 49 / 664 24 90 18
- Correo electrónico: asier.calzada@ehu.eus

DATOS RELATIVOS AL PROYECTO:

- **Título del proyecto de tesis**: "La dictoglosia colaborativa en el aula de inglés como lengua extranjera en Educación Primaria y sus efectos en la adquisición de la gramática del inglés"
- **Financiado por el Ministerio de Educación**
- Descripción del proyecto:
  - Este estudio tiene como objetivo explorar el potencial de la escritura colaborativa en el aprendizaje del inglés como lengua extranjera y documentar la interacción colaborativa entre el alumnado de Educación Primaria mientras realizan tareas que requieran la elaboración de un producto escrito.

DESCRIPCIÓN DEL PROCEDIMIENTO

- **Tipo de procedimiento**: los participantes completarán pruebas escritas en inglés y castellano, cuestionarios de perfil lingüístico y de actitud y una tarea de producción escrita (dictoglosia).
  - La tarea de escritura podrá ser colaborativa o individual.
  - Los participantes escucharán una breve historia dos veces y, tras ello, deberán reconstruir la historia, que deberá guardar similitud con la original.
  - Antes y después de la tarea se les pedirá su opinión sobre aspectos relacionados con la escritura en lengua extranjera, el trabajo colaborativo y su experiencia en el aprendizaje de la lengua extranjera en general.
  - Datos personales anónimos: los datos personales serán tratados de forma totalmente anónima, así como los resultados de todas las pruebas.
- **Número de intervenciones:** se necesitarán 8 sesiones distribuidas a lo largo de 3 semanas en **(MONTH)** (6 horas en total). Los cuestionarios y las pruebas escritas se recogerán en seis sesiones distintas (4 h en total). Las tareas de escritura se realizarán en dos sesiones de 1 hora de duración aproximada (2 h en total). Todas las pruebas y tareas se llevarán a cabo en **(NAME OF THE SCHOOL)**. En caso de que usted no conceda el permiso de que su hijo/a colabore en este proyecto, debido a la cantidad de tiempo que conlleva la participación, están pensadas una serie de actividades de repaso para realizar en el aula con su profesor/a.
- **Descripción del procedimiento:** Los diferentes tests de nivel y cuestionarios se realizarán de manera individual. Para la tarea de escritura, en cada sesión el/la participante completará una tarea

1

e

bien colaborativa (en parejas), junto a un compañero/a, o bien de forma individual. Estas tareas tienen como objeto obtener información sobre los beneficios de las tareas colaborativas en la producción escrita de los alumnos de Educación Primaria.

- **Descripción de riesgos:** no existe ningún riesgo para el/la alumno/a.

DERECHOS DEL PARTICIPANTE:

- La participación en este estudio es **voluntaria** y podrá dejar de participar en cualquier momento, sin que ello suponga ningún perjuicio, comunicando la intención de abandono al investigador mediante correo electrónico.
- Si usted concede el permiso de que su hijo/a colabore en este proyecto, una vez haya finalizado, tendrá a su **disposición** toda la información relativa a los resultados obtenidos en el mismo, respetando la confidencialidad de los participantes. Puede obtener los datos poniéndose en contacto con el investigador.
- **Las pruebas también pueden incluir la recogida de información mediante grabaciones:**
    - **O Doy el consentimiento para la grabación**
    - **O NO doy el consentimiento para la grabación**

- Los datos personales que nos ha facilitado para este proyecto de investigación serán tratados con absoluta **confidencialidad** de acuerdo con la Ley de Protección de Datos. Los datos personales recabados objeto de esta investigación serán incluidos en un fichero registrado en la Agencia Vasca de Protección de Datos (nº **INM 0089)** bajo el nombre **Collaborative writing and young learners**. El responsable del fichero o tratamiento será la UPV/EHU.
  Puede consultar en cualquier momento los datos que nos ha facilitado o solicitarnos que rectifiquemos o cancelemos sus datos o simplemente que no los utilicemos para algún fin concreto de esta investigación. La manera de hacerlo es dirigiéndose al Responsable de Seguridad LOPD de la UPV/EHU (Rectorado, Barrio Sarriena s/n, 48940 Leioa-Bizkaia).
  Para más información sobre Protección de Datos le recomendamos consultar en Internet nuestra página web www.ehu.es/babestu.

IDENTIFICACIÓN DE LA PERSONA QUE PRESTA EL CONSENTIMIENTO

Yo (nombre y apellidos) ……………………………………………. con D.N.I. …………………..

como padre / madre / representante legal de (nombre y apellidos del alumno/a) ……………………………………………

MANIFIESTO

que he entendido que este consentimiento puede ser revocado por mí en cualquier momento y **OTORGO MI CONSENTIMIENTO** para participar en este estudio.

(Fecha) (Firma del padre/ madre / representante legal)

A17 Oral interaction transcription sample

@Begin

@Language: eng, eus, spa

@School: mariturri

@Researcher: Asier

@Participants: *CHI A MAR009, *CHI B MAR012

*CHI B: <tengo muy pocas cosas tengo>@spa! Laura lunch at school forgot eh taking the chocolate of peanuts <pero es que tiene chocolate negro>@spa.

*CHI A: <era negro el chocolate>@spa a dark chocolate.

*CHI B: <se come el chocolate negro con xxx>@spa she is hungry at school is eh she eat the chocolate is good she go to class she feel sick and her face is becoming red forgot she is allergic for peanuts and go to the doctor at the next day <eso es todo>@spa.

*CHI A: <lo tienes todo madre mia>@spa.

*CHI B: <no tengo gran cosa si lees lo he puesto en indio te lo voy a leer eh>@spa? Laura lunchchh forgot a mu chocolate peanuts go class sick forgot allergic for peanuts doctor [laughs].

*CHI A: <bueno pero lo has entendido yo solo he entendido>@spa howich supermarket apple eh chocolate…

*CHI B: <has puesto>@spa superapple?

*CHI A: <si porque no me daba tiempo para mas datos>@spa chocolate chocolate in class <le da alergia y tiene matematicas>@spa…

*CHI B: +<en clase no en el patio>@spa! <pero como se llamaba patio>@spa?

*CHI A: eh… playground.

*CHI B: <voy a escribirlo>@spa.

*RES: CHI B speak up! as if you were in class.

*CHI B: <pero en clase no nos estan grabando>@spa…

*RES: <venga venga>@spa.

*CHI B: <como se llamaba el titulo era no se que no se que Laura o lunch Laura>@spa…

*CHI A: +<alergia a los cacahuetes>@spa allergic to… Laura's allergy…

*CHI B: allergic Laura <creo que era>@spa…

*CHI A: <tu ponlo>@spa!

*CHI B: <no hacia falta titulo aqui no pone titulo>@spa [pointing at the the reconstruction photocopy].

*CHI A: <hala pues ya el principio Laura tiene que no se que no se cuantos y su padre>@spa…

*CHI B: Laura have to go to school but her dad her father haven't her father forgot eh <como es almuerzo>@spa? eh… lunch her lunch!

*CHI A: <no lunch es comida>@spa.

*CHI B: <ya>@spa… Asier how do you say almuerzo? lunch? is lunch?

[RES gestures but is not visible in the video]

*CHI B: <hala pues>@spa lunch and she go to the <tienda>@spa to buy chocolate.

*CHI A: <no eh prepara un sandwich eso es lo que me hice yo>@spa.

*CHI B: <un sandwich>@spa? <no pero si ella va a la escuela no tiene tampoco chocolate>@spa?

*CHI A: <si pero prepara un sandwich el padre y luego ella se va a la tienda a por chocolate>@spa.

*CHI B: <no a su padre se le olvida preparar el almuerzo y ella se va a por chocolate>@spa.

*CHI A: <pues se le olvida prepar el sandwich>@spa.

*CHI B: <pues el sandwich (,) pero a mi se ha olvidado preparar una pizza tampoco sera mas>@spa!

*CHI A: <a ver Laura tiene que ir al cole>@spa.

*CHI B: Laura have to go to school… <vale aqui llega mi gran pregunta siempre se me olvida como se escribe school (,) escribo siempre mal ese coool >@spa.

*CHI A: <ese creo que no (,) creo que es ese hache ce>@spa.

*CHI B: hee [meaning agreement yes].

*CHI A: <tiene mas sentido>@spa.

*CHI B: <sino seria>@spa to ↑shool.

*CHI A: shool haha <es Laura o Lola>@spa?

*CHI B: Lau… Laura <no es Lola Lola es perrito>@spa <has traido el boli borrable>@spa?

*CHI A: <no tengo boli borrable>@spa.

*CHI B: Laura have to go to school but her father eh…

*CHI A: don't prepare…

*CHI B: didn't? or forgot? or but her father…

*CHI A: +<but her father…

*CHI B: forgot.

*CHI A: don't or <como se llamaba>@spa?

*CHI B: <he dicho>@spa forgot <se ha olvidado>@spa.

*CHI A: or forgots prepare her sandwich.

*CHI B: [repeating while writing] her father…

*CHI A: fotarrr faterrr fotarrr.

*CHI B: forgot… her lunch.

*CHI A: her sandwich! <es que si no no esta bien>@spa [taking an object from the table] <oye quieres con ser ya no te pinchas>@spa.

*CHI B: <solo digo una cosa si escribo mal sandwich es tu problema>@spa.

*CHI A: sandwich.

*CHI B: <no pero es que>@spa sand-wich <asi>@spa?

*CHI A: <no no tiene hache>@spa.

*CHI B: <ah no>@spa?

*CHI A: <no>@spa.

*CHI B: <si si tiene>@spa.

*CHI A: <no no tiene>@spa.

*CHI B: <escribelo tu que no se como se escribe>@spa.

*CHI A: <pues como lo has puesto pero sin hache>@spa.

*CHI B: <sin hache>@spa like this?

*CHI A: <creo que si>@spa.

*CHI B: <vale>@spa sandwich <suena rarito si lo lees>@spa Laura have to go to sssscol&(school) [pronouncing in spa] but her father forgot her sandwich [/] sandwich [laughs] <vale>@spa.

*CHI A: Lola go to the supermarket…

*CHI B: +<she takes money…

*CHI A: money money money.

*CHI B: and go to the supermarket.

*CHI A: to buy…

*CHI B: chocolate.

*CHI A: chocolate <tiene algo por detras>@spa chocolate…

*CHI B: peanuts chocolate.

*CHI A: chocolate peanuts <sera>@spa.

*CHI B: chocolate with peanuts <es chocolate con> peanuts.

*CHI A: <vale tu eres la que entiende>@spa.

*CHI B: she takes the money…

*CHI A: <yo soy el que utiliza la cabeza>@spa…

*CHI A: she takes takessss [exaggerating the 3rd s] <vale>@spa.

*RES: [to the whole class] have a look at the drawing it can give you a clue.

*CHI A: <no a los cacahuetes>@spa!

*CHI B: <para maxim maxim podia comer de lo mio porque no llevaba avellanas y no puede comer avellanas>@spa and go…

*CHI A: and buy and go to buy… no.

*CHI B: and go to buy…

*CHI A: to the supermarket.

*CHI B: and go to the supermarket… super… market…

*CHI A: <como seria>@spa? to buy <o que habria que poner>@spa?

*CHI B: to buy…

*CHI A: <otro>@spa to? and go to the supermarket to buy <bueno puede ser>@spa.

*CHI B: to buy… chocolate with peanuts… chocolate <se escribe igual>@spa?

*CHI A: <si creo que si (,) espera se puede mirar en el texto de atras esta escrito>@spa.

*CHI B: peanuts.

*CHI A: <chocolate si se escribe igual>@spa.

*CHI B: and she go to school no?

*CHI A: <si>@spa.

*CHI B: and she go…

*CHI A: to school.

*CHI B: <hala he puesto sool>@spa she goes to school… <creo que lo he vuelto a escribir mal>@spa.

*CHI A: <si si has escrito mal si escribir mal>@spa.

*CHI B: [pronouncing in spa] school…. [pronouncing in spa] goes (…) in the playground when she is hungry…

*CHI A: <no primero tienes que decir que tenia matematicas>@spa.

*CHI B: <no pero eso es despues en el patio lo come>@spa and <luego tenia luego>@spa go hungry school class <o sea primero se lo come y luego se va a clase de mate se siente como mala y con granos>@spa.

*CHI A: <y va al doctor y muere>@spa!

*CHI B: <no>@spa! <es una historia y final feliz>@spa.

*CHI A: <pues>@spa go to playground.

*CHI B: she go to playground…

*CHI A: <yo que pondria algun punto>@spa!

*CHI B: <algun punto>@spa? Laura have to go to school comma but her father forgot her sandwich <punto>@spa she takes money and go to the supermarket to… to…

*CHI A: buy!

*CHI B: <ostras se me ha olvidado>@spa!

*CHI A: <te has comido el>@spa buy!

*CHI B: to buy chocolate with peanuts and she goes to the school <punto>@spa she s… when she go to the classroom she eat she is hungry and she eat the chocolate bar.

*CHI A: she is hungry and she eat her or his her or his (,) his his!

*CHI B: <no es>@spa her.

*CHI A: his.

*CHI B: <es>@spa her (,)spa his <es para chicos>@spa…

*CHI A: <tengo un lio ahi>@spa…

*CHI B: <a ver>@spa her <para chicas>@spa <y>@spa his <para chicos>@spa (,) <yo antes ponia his para todo>@spa when… when she go… to the playground… they… <esto esta bien escrito no>@spa?

*CHI A: <el que>@spa?

*CHI B: playground <si no>@spa?

*CHI A: <si>@spa.

*CHI B: she is very hungry… no she ↑was very hungry very <siempre me equivoco>@spa very <es con be o con uve>@spa? <es con be no>@spa?

*CHI A: <yo creo que es con uve>@spa?

*CHI B: <pues con uve>@spa [pronounced in spa] hungry <pero o sea tambien me equivoco>@spa angry <y>@spa hungry.

*CHI A: <es tambien con hache uno con hache>@spa (…) <tu piensa angry birds no son pajaros hambrientos>@spa.

*CHI B: [laughs] angry birrrds…

*RES: how are you doing?

*CHI B: fine! with <fallos ortograficos pero bueno>@spa she is hungry and she eat…

*CHI A: chocolate with… <ya se me ha olvidado>@spa…

*CHI B: the chocolate with peanuts.

*CHI A: pinotes.

*CHI B: chocolate.

*CHI A: with pinotes.

*CHI B: with… pinut (…) she go to mats&(maths) class?

*CHI A: <punto>@spa later.

*CHI B: later she goes…

*CHI A: [pronouncing in spa] later.

*CHI B: <como los guantes de late no lo olvideis no hagais esto en cas niños>@spa.

*CHI A: later she go…

*CHI B: to maths class.

*CHI A: she go to class and he's got maths.

*CHI B: class and she's got mats (…) she's got class <iba a poner>@spa she's got maths…

*CHI A: <hombre si va a clase no va a tener clase y>@spa class <es clase se escribe clase no ves que lo tienes ahi [pointing at the top of the photocopy]>@spa.

*CHI B: <no (,) esta en castellano>@spa.

*CHI A: <ah ya decia yo>@spa.

*CHI B: <ay no se nos ha olvidado que ha dicho que el chocolate estaba bueno pero da igual da igual eso sobre que el chocolate estaba bueno xxx que chocolate negro>@spa (…) <ah que iba a decir si te has dado cuenta los que menos nivel tienen de ingles en toda la clase se han quedado (…) tienen menos nivel de ingles porque no van a academia bueno tu tampoco vas a academia pero tu me copias la mitad>@spa.

*CHI A: <no eres tu la que me copias a mi>@spa.

*CHI B: <no>@spa…

[CHI and CHI B talk about class things for 15 s]

*CHI A: <a ver>@spa she is good in the class he… she's feeling…

*CHI B: sick and her face it… <como se esta poniendo>@spa… it's red more red <a ver>@spa in the class… feeling… [pronouncing in spa] feeling… <creo que esto se escribe asi>@spa.

*CHI A: feeling <si>@spa.

*RES: finished here?

*CHI B: no!

*CHI A: <casi casi>@spa.

*RES: <venga>@spa almost.

*CHI B: she's feeling… she…

*CHI A: <tenemos diez minutos>@spa.

*CHI B: and her face is traumatic… [laughs] is traumatic… is red… <rojo>@spa and then says they went to the doctor… no (,) red she forgot [/] she was forgot… was allergic for peanuts…

*CHI A: <si porque al chocolate no>@spa!

*CHI B: she forget…

*CHI A: xxx.

*CHI B: she was allergic…

*CHI A: for or to? to the peanuts or for?

*CHI B: for the peanuts (,) no (,) to the peanuts allergic for the peanuts or allergic to the peanuts… to the peanuts.

*CHI A: <a los pinos o por los pinos (,) por los pinos>@spa?

*CHI B: to the peanuts.

*CHI A: <por los pinos y las pinas este es mi caso>@spa.

*CHI B: she allergic to the peanuts.

*CHI A: [pronouncing in spa] pinotes.

*CHI B: peanuts eh… <punto>@spa the next day she go to the doctor and <ya esta>@spa… and the next day…

*CHI A: [pronouncing in spa] day [/] day… she [exaggerating the dipthong] she go [corrects himself] she goes!

*CHI B: goes to the doctor…

*CHI A: [pronouncing in spa] doctore eh doctore.

*CHI B: <doktore doktore>@eus.

*CHI A: doctor.

*CHI B: <ya esta que no se me ocurre nada mas>@spa Laura have to go to the school and her father forgot her sandwich she takes money and go to the supermarket to buy chocolate with peanuts and she goes to the school… to ↑the school.

*CHI A: <joe chaval tu te saltas cosas que te las comes>@spa.

*CHI B: <tu no escribes asi que calla>@spa when she goes to playground she was very hungry… she was very ↑hun gry and she [/] and she eat the chocolate with peanuts later she go to class and she's got maths in the class she is feeling sick and her face is red she forgot she is allergic to the peanuts! at the next day on the next day she goes [/] she goes to the doctor <ya esta>@spa!

*CHI A: finish!

*CHI B: Asier we've got finish.

*RES: did you revise the text? <habeis leido en alto una vez>@spa?

*CHI B: <si>@spa.

*RES: ok good thank you!

@End

A18 Interrater guidelines for assessing LREs

May 2020 Asier Calzada PhD dissertation

Interrater analysis guidelines

1. LREs

20 % Dataset

LREs are classified according to this classification (examples from other dyad interactions are provided for each category, i.e. excluded from interrater analysis)

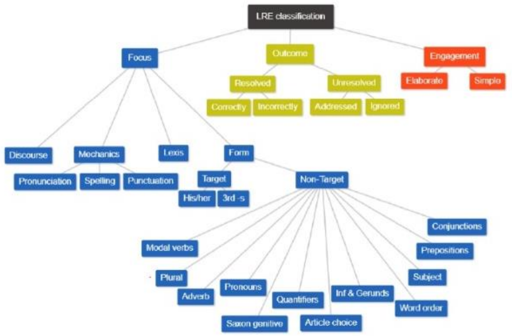


Figure 1 Overview of the LRE classification

- **Focus**
  - Discourse

    **CHI B: eh... the title eh I don't...*

    **CHI A: ¡me tienes que decir a mí!*

    **CHI B: ya lo sé que no me acuerdo del título.*

    **CHI A: ah, ¡que tiene título! [giggles].*

  - Form
    - Non-target

      1. Adverb

      **CHI A: after o...? ¿pero aquí after? after she eat... the peanus&(peanuts), sí, pon, after eat the peanos&(peanuts)...*

      **CHI B: after thing?*

1

May 2020 Asier Calzada PhD dissertation

**CHI A: sí, she eat the peanout&(peanut) and remember que ella es alérgica (...) and remember... she is... allergic to allergic to peanoux&(peanuts) and and her dad go to hospital with Laura.*

2. Article choice

**CHI A:*	*go to hospital.*
**CHI B:*	*to a hospital, sería mejor.*
**CHI A:*	*[whispering] da igual.*

3. Conjunctions

**CHI A:*	*eh Laura...*
**CHI B:*	*but pon 'but'.*
**CHI A:*	*no, porque...*
**CHI B:*	*no tenían no tenían chocolate tenían que ir a comprar chocolate, but...*
**CHI A:*	*but.*

4. Determiners

**CHI A:*	*sandwich.*
**CHI B:*	*one sandwich with...*

5. Infinitives & Gerunds

**CHI B:*	*putting...*
**CHI A:*	*putting...*
**CHI B:*	*hahaha, bueno, putting...*
**CHI A:*	*bueno, put...*
**CHI B:*	*red but she start eh... [they laugh again] putting.*
**CHI A:*	*putting to put red.*

6. Modal verbs

**CHI A: the Laura she can't's she eats she eat eh... peanuts and she can't*

7. Object

**CHI B:*	*was playing with her...*
**CHI A:*	*uncle, no, was playing football...*

2

8. Plural

CHI A: and you put [reading CHI B's notes] uncle dad put funny face...

**CHI B: funny faces.*

9. Preposition

**CHI B: but eh... in the supermarket.*

**CHI A: on the supermarket.*

**CHI B: on, sí.*

10. Pronouns

**CHI B: we prepare.*

**CHI A: no, their prepare.*

**CHI B: their prepare...*

11. Quantifiers

**CHI B: Maria is show...*

**CHI A: to bottle is to bottle.*

**CHI B: one bottle.*

12. Saxon Genitive

**CHI B: aquí pon, eh Laura...*

**CHI A: no, the father los padres de Laura.*

**CHI B: the Laura's father...*

**CHI A: ¿pongo esto?*

**CHI B: no, Laura's father.*

13. Subject

**CHI A: in the afternoon is very... no, the afternoon is very good and...*

14. Tense

**CHI A: eh, a ver es que... (...) Lara they go went to school*

15. Verb

**CHI B: pues vamos a poner aquí, wonderful day.*

**CHI A: it is a wonderful day... it is a... sí, vamos a escribir todo eso.*

16. Word order

**CHI B: es que, they say... chocolate black.*

**CHI A: black chocolate.*

**CHI B: black chocolate, eh...*

- Target

1. 3 -s

**CHI B: Lola I love chocolate Lola loves chocolate prr [as if she is not sure].*

2. His/her

**CHI B: but her father but.*

**CHI A: his father.*

**CHI B: but her father no but bat&(but).*

- <u>Lexis</u>

(I) Looking for English equivalent

**CHI B: ¿cómo se decía 'coger'?*

**CHI A: take the money.*

(II) Looking for alternatives in English

**CHI B: eh we can put her father today.*

**CHI A: one day.*

- <u>Mechanics</u>
  - Pronunciation

**CHI A: apel&(apple) [pronounced correctly] apple*

  - Punctuation

**CHI A: she xxx xxx chocolate xxx chocolate no me cabe, late she.*

**CHI B: comma.*

  - Spelling

**CHI A: [Spanish pronunciation spelling] note books?*

**CHI B: sí, con dos os.*

**CHI B: eh, notebooks.*

• **Outcome**

- Resolved (only those LREs that are incorporated in the writing)
  - Correctly
  - Incorrectly
- Unresolved (LREs which are not incorporated in the writing or which the learners start discussing but finally the researcher resolves)
  - Addressed

(I)

**CHI A:* *with class mat&(maths), no mate class, algo así.*

**CHI B:* *se escribe así.*

**CHI A:* *bah, qué más da.*

(II)

**CHI A:* *and the and the doctor go to Laura…*

**CHI B:* *go to Laura go to with Laura go with Laura.*

**CHI A:* *to with Laura.*

**CHI B:* *with Laura fue con Laura donde Laura, bueno da igual.*

- Ignored

(I) including self-talk which is not replied by the peer

**CHI A:* *of Laura and sh… he gives money to buy a… no.*

(II) or questions directly enquired to their peer which are left unanswered

**CHI A: y después, ¿cómo se dice? and [/] and [/] and… bueno ahora te toca.*

• **Engagement (McDonough & Hernández González, 2019)**

- Simple: including self-corrections and instances where the peer answers either repeating the solution provided by the LRE initiator or answers with a simple confirmation ("ok", "vale") or with a phatic utterance ("ah", "uhmm").
- Elaborate: showing an active participation from both members of the dyad

2. Individual languaging

Individual languaging is **only** classified in terms of **FOCUS** (see list above)

100 % Dataset (excel file, two tabs: Dictogloss 1 and Dictogloss 2)

3. Use of Previously Known Languages (PKL) in LRE turns

Criteria: are there more words in L2/PKL in **each turn**? (Storch and Aldosari, 2013)

Same 20 % as in LREs

4. Pair dynamics/Patterns of interaction (Butler & Zeng, 2015; García Mayo & Imaz Agirre, 2019; Storch, 2002)

Same 20 % as in LREs

A19 Example of handwritten production

LASLAB

Universidad del País Vasco Euskal Herriko Unibertsitatea

T1

1) Nombre y apellidos: ABE059	Clase: XXX
2) Nombre y apellidos: ABE058	Clase: XXX

Ahora tenéis que escribir el texto. Recordad que debe ser lo más parecido posible al que acabáis de escuchar.

Now you have to write your text. Remember that it should be as similar as possible to the one you have just listened to.

The father of Lora, forgets prepare the sandwich.
He gives a money for buy enething in the supermarket.
Lora is hungry she go to the supermarket and buys
a chocolate-bar with peanuts.
She ate chocolate-bar and she go to the class, and
She has maths. The teacher, say to Lora she has
a red points in the pace.
Lora rember that she don't can eat peanuts. The
teacher cal to the father of Lora, She eat
peanuts. The father of Lora go to the
school and take Lora and then go together
to the hospital. THE END

THANK YOU FOR YOUR TIME!

A20 DG1 and DG2 Idea Unit abstraction

The original texts have been divided into 14 Idea Units each. The coder should use lax criteria to interpret that a certain idea unit has been retrieved by the learner. That is, do not have a "word by word" idea in mind. An abstraction for each idea unit has been added to help the coder decide.

	Naughty Laura	
1.	Laura **takes** her lunch box everyday to school.	Getting ready for school
2.	Today, her father forgot to prepare a sandwich,	Forgetting food
3.	so he **gives** her some money	Father gives money
4.	to buy an apple at the supermarket.	Buying at the supermarket
5.	At the supermarket, she **sees** some chocolate bars.	Chocolate at the supermarket
6.	She **loves** chocolate,	Loving chocolate
7.	so she **buys** one with black chocolate and peanuts instead of an apple.	Buying peanuts
8.	At the break she **feels** very hungry and **eats** the chocolate bar.	Eating at school
9.	It **tastes** so good!	Good taste
10.	Then, she **returns** to class. They have Maths.	Back to class
11.	Suddenly, her face **turns** red and she **starts** to feel very sick.	Feeling sick
12.	Laura forgot she is allergic to peanuts!	Allergy
13.	The teacher **calls** her father	Teacher & Father
14.	and he **drives** her to hospital.	Hospital

A celebration

1.	The Smiths are spending a day together in the garden.	Family in the garden
2.	Tom is playing football with **his** uncle.	Tom uncle football
3.	He calls **his** grandmother to join them, but she is busy playing cards and she is winning all the time!	Grandma cards
4.	Now it is lunchtime.	Lunchtime
5.	They are all sitting at the table in the garden.	Table garden
6.	María, the oldest granddaughter, asks **her** mum to take some pictures.	Pictures
7.	**Her** dad pulls funny faces and they laugh.	Dad funny
8.	In the afternoon, it’s very hot.	Hot
9.	Tom sees that **his** grandfather is preparing ice cream in the kitchen.	Ice cream
10.	He calls **his** sister and they go quickly to try it.	Trying ice cream
11.	It’s delicious!	Good taste
12.	María gives a big hug to **her** grandfather.	Maria gives grandfather something
13.	Then, she takes a portion for **her** aunt, who is in the swimming pool.	Aunt swimming pool ice cream
14.	What a wonderful day!	Good day

A21 L2 writing assessment rubric for the dictogloss tasks (based on Government of Navarre, 2017)

		3	2	1
TASK	**Adequacy**	Most ideas from the text are mentioned, all the parts of the story are included (beginning, body, ending); the length of the text is appropriate	Just some ideas from the text are mentioned, most parts of a story are included; the text is too short	Notable omissions of the content points and/or considerable irrelevance of some of them
	Coherence	A clear text, easy to understand	Easy to understand, although there are some incoherent points that confuse the reader	Difficult to understand
LANGUAGE	**Cohesion**	Ideas are well organized (use of paragraphs). Cohesive devices linking sentences and paragraphs. No serious mistakes	Ideas are organized. Some cohesive devices linking sentences and paragraphs. There may be some mistakes	There is a lack of organization or linking devices
	Grammatical accuracy	Very few, irrelevant or no grammar errors at all. Good command of grammar	Some acceptable grammar errors. Fair command of English grammar	Serious and numerous grammar mistakes
	Mechanics	Most words are written correctly, only some occasional mistakes	Some spelling mistakes (between 3 and 6), some of them in basic vocabulary	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

A22 L2 writing assessment rubric for the continuation task (based on Government of Navarre, 2017)

		3	2	1
TASK	Adequacy	All the parts of the story are included (beginning, body, ending); the length of the text is appropriate	Most parts of a story are included; the text is too short (ideas are not fully developed)	Notable omissions of the content and/or considerable irrelevance of some of them
	Coherence	A clear text, easy to understand	Easy to understand, although there are some incoherent points that confuse the reader	Difficult to understand
LANGUAGE	**Cohesion**	Ideas are well organized (use of paragraphs). Cohesive devices linking sentences and paragraphs. No serious mistakes	Ideas are organized. Some cohesive devices linking sentences and paragraphs. There may be some mistakes	There is a lack of organization or linking devices
	Grammatical accuracy	Very few, irrelevant or no grammar errors at all. Good command of grammar	Some acceptable grammar errors. Fair command of English grammar	Serious and numerous grammar mistakes
	Mechanics	Most words are written correctly, only some occasional mistakes	Some spelling mistakes (between 3 and 6), some of them in basic vocabulary	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

A23 L1 writing assessment rubric (based on Fernández et al. 2019)

CRITERIA	LEVELS OF ACHIEVEMENT
1. Topic or Inciting Incident	1. What happened at the beginning is not explained. 2. What happened at the beginning is partially explained, but it is not comprehensible. 3. What happened at the beginning is partially explained, as well as how the characters felt. 4. What happened at the beginning is clearly explained, as well as how the characters felt and what they planned to do.
2. Plot (episodes and resolution)	1. Nothing happens in the story or what happens is not comprehensible. 2. The story is a bit confusing or some facts do not make much sense 3. The story is simple. What happened next and at the end is explained with enough clarity. 4. Although the story is long, what happened is very clear (in well-organized segments), as well as what happened to the characters at the end.
3. Creativity and Interest	1. The story is copied or cannot be understood. 2. The story is original and can be understood, but it is too short or very boring. 3. The story is original and entertaining. 4. The story is original and entertaining. It is nice, shows a sense of humor, or teaches us something.
4. Sentences	1. Most sentences cannot be understood because they are badly built. 2. Most sentences can be understood, but some are badly built. 3. The sentences are easily understood, but there are too few punctuation marks (or there are only commas). 4. The sentences are easily understood, most of the punctuation marks are in the right places, even in the dialogues.
6. Vocabulary	1. The vocabulary is very poor and with many mistakes (some sentences are not correctly representing what they are trying to say). 2. The words are correct, but there are a lot of repetitions. 3. The vocabulary used is correct and varied (not repeated too often). 4. The vocabulary used is rich, with some little-known words.
7. Spelling	1. There are a lot of spelling mistakes (at least one in each sentence). 2. There are quite a lot of spelling mistakes (at least one in every other sentence). 3. There are few spelling mistakes. 4. There are no spelling mistakes.

APPENDIX B

B1 Inferential statistics for RQ1a and RQ1b

Paired-samples T-tests for LREs outcome on Day 1 / Focus

Measure 1		Measure 2	t	df	*p*	Cohen's *d*
FTARGETS_RES_COR	-	FTARGETS_RES_INC	2.457	28	.020	0.456
FTARGETS_RES_COR	-	FTARGETS_UNRES_ADD	2.608	28	.014	0.484
FTARGETS_RES_COR	-	FTARGETS_UNRES_IG	2.196	28	.037	0.408
FTARGETS_RES_INC	-	DISC_UNRES_ADD	1.307	28	.202	0.243
FTARGETS_RES_INC	-	FTARGETS_UNRES_IG	0.626	28	.537	0.116
FTARGETS_UNRES_ADD	-	FTARGETS_UNRES_IG	-1.361	28	.184	-0.253

Measure 1		Measure 2	t	df	*p*	Cohen's *d*
FOTHERS_RES_COR	-	FOTHERS_RES_INC	2.872	28	.008	0.533
FOTHERS_RES_COR	-	FOTHERS_UNRES_ADD	4.998	28	< .001	0.928
FOTHERS_RES_COR	-	FOTHERS_UNRES_IG	3.547	28	.001	0.659
FOTHERS_RES_INC	-	FOTHERS_UNRES_ADD	4.126	28	< .001	0.766
FOTHERS_RES_INC	-	FOTHERS_UNRES_IG	0.955	28	.348	0.177
FOTHERS_UNRES_ADD	-	FOTHERS_UNRES_IG	-2.822	28	.009	-0.524

Measure 1		Measure 2	t	df	*p*	Cohen's *d*
LEX_RES_COR	-	LEX_RES_INC	4.732	28	< .001	0.879
LEX_RES_COR	-	FOTHERS_UNRES_ADD	5.137	28	< .001	0.954
LEX_RES_COR	-	LEX_UNRES_IG	3.780	28	< .001	0.702
LEX_RES_INC	-	LEX_UNRES_ADD	-2.714	28	.011	-0.504
LEX_RES_INC	-	LEX_UNRES_IG	-1.072	28	.293	-0.199
LEX_UNRES_ADD	-	LEX_UNRES_IG	1.944	28	.062	0.361

Measure 1		Measure 2	t	df	*p*	Cohen's *d*
MECH_RES_COR	-	MECH_RES_INC	7.893	28	< .001	1.466
MECH_RES_COR	-	MECH_UNRES_ADD	8.219	28	< .001	1.526
MECH_RES_COR	-	MECH_UNRES_IG	8.308	28	< .001	1.543
MECH_RES_INC	-	MECH_UNRES_ADD	3.550	28	.001	0.659
MECH_RES_INC	-	MECH_UNRES_IG	2.807	28	.009	0.521
MECH_UNRES_ADD	-	MECH_UNRES_IG	-0.441	28	.663	-0.082

Measure 1	Measure 2	t	df	*p*	Cohen's *d*
DISC_RES_COR	- DISC_RES_INC	1.548	28	.133	0.287
DISC_RES_COR	- DISC_UNRES_ADD	2.635	28	.014	0.489
DISC_RES_COR	- DISC_UNRES_IG	1.294	28	.206	0.240
DISC_RES_INC	- DISC_UNRES_ADD	1.000	28	.326	0.186
DISC_RES_INC	- DISC_UNRES_IG	-0.441	28	.663	-0.082
DISC_UNRES_ADD	- DISC_UNRES_IG	-1.307	28	.202	-0.243

Paired-samples T-tests for LREs engagement on Day 1

Measure 1	Measure 2	t	df	*p*	Cohen's *d*
FTARGETS_ELA	- FTARGETS_SIM	-0.708	28	.485	-0.131
FOTHERS_ELA	- FOTHERS_SIM	-1.901	28	.068	-0.353
LEX_ELA	- LEX_SIM	3.852	28	< .001	0.715
MECH_ELA	- MECH_SIM	-0.691	28	.495	-0.128
DISC_ELA	- DISC_SIM	-0.550	28	.586	-0.102

Paired-samples T-tests for LREs PKL and L2 use on Day 1

Measure 1	Measure 2	t	df	*p*	Cohen's *d*
TARGET_TOTAL_L1	- TARGET_TOTAL_L2	-0.584	28	.564	-0.108
FOTHERS_L1	- FOTHERS_L2	-3.606	28	.001	-0.670
LL1	- LL2	1.299	28	.205	0.241
ML1	- ML2	3.967	28	< .001	0.737
DL1	- DL2	2.012	28	.054	0.374

Paired-samples T-tests for LREs PKL and L2 turn length (in words) on Day 1

Measure 1	Measure 2	t	df	*p*	Cohen's *d*
TARGET_ws_L1	- TARGET_ws_L2	0.324	28	.749	0.060
FOTHERSL1_ws	- FOTHERSL2_ws	-1.743	28	.092	-0.324
LL1_ws	- LL2_ws	1.621	28	.116	0.301
ML1_ws	- ML2_ws	3.185	28	.004	0.591
DL1_ws	- DL2_ws	2.467	28	.020	0.458

B2 Inferential statistics for RQ1c

WLREs in Comp

Measure 1		Measure 2	Statistic	df	*p*	Cohen's *d*
Disc_day1	-	Disc_day2	0.572	25	.571	0.098
Fothers_day1	-	Fothers_day2	1.000	25	.327	0.196
Lexis_day1	-	Lexis_day2	1.494	25	.148	0.293
Mech_day1	-	Mech_day2	- 0.296	25	.770	-0.058

LREs in Collab and FFI+Collab

Repeated measures ANOVA on Time on Task

Cases	Sum of Squares	df	Mean Square	F	*p*	η^2
Time	247392.566	1	247392.566	4.895	.036	0.040
Time ✻ ExpGroup	286989.255	1	286989.255	5.678	.024	0.046
Residuals	1.365e +6	27	50539.901			

Cases	Sum of Squares	df	Mean Square	F	*p*	η^2
ExpGroup	2.369e +6	1	2.369e +6	33.450	< .001	0.383
Residuals	1.912e +6	27	70810.592			

Post Hoc Comparisons - Time

			95% CI for Mean Difference					
		Mean Diff	Lower	Upper	SE	t	Cohen's *d*	*p* bonf
Day.1	Day.2	130.698	9.489	251.906	59.073	2.212	0.411	.036

Post Hoc Comparisons - ExpGroup

			95% CI for Mean Difference					
		Mean Diff	Lower	Upper	SE	t	Cohen's *d*	*p* bonf
Collab	FFI+Collab	-404.407	-547.878	-260.936	69.924	-5.784	-1.074	< .001

Post Hoc Comparisons - ExpGroup ✻ Time

		Mean Diff	95% CI for Mean Difference Lower	Upper	SE	t	p_{bonf}
Collab, Day.1	FFI+Collab, Day.1	-545.176	-796.156	-294.196	91.537	-5.956	< .001
	Collab, Day.2	-10.071	-251.974	231.831	84.971	-0.119	1.000
	FFI+Collab, Day.2	-273.710	-524.689	-22.730	91.537	-2.990	.025
FFI+Collab, Day.1	Collab, Day.2	535.105	284.125	786.085	91.537	5.846	< .001
	FFI+Collab, Day.2	271.467	37.767	505.167	82.089	3.307	.016
Collab, Day.2	FFI+Collab, Day.2	-263.638	-514.618	-12.658	91.537	-2.880	.034

Repeated measures ANOVA on LRE Focus

F-targets

Cases	Sum of Squares	df	Mean Square	F	*p*	η²
Time	1.564	1	1.564	0.659	.424	0.007
Time ✻ ExpGroup	3.219	1	3.219	1.357	.254	0.015
Residuals	64.057	27	2.372			

Cases	Sum of Squares	df	Mean Square	F	*p*	η²
ExpGroup	14.900	1	14.900	3.082	.090	0.070
Residuals	130.514	27	4.834			

F-others

Cases	Sum of Squares	df	Mean Square	F	*p*	η²
Time	5.214	1	5.214	0.919	.346	0.008
Time ✻ ExpGroup	2.317	1	2.317	0.408	.528	0.003
Residuals	153.200	27	5.674			

Cases	Sum of Squares	df	Mean Square	F	*p*	η²
ExpGroup	4.968	1	4.968	0.263	.612	0.007
Residuals	509.514	27	18.871			

Lexis

Cases	Sum of Squares	df	Mean Square	F	*p*	η²
Time	1.105	1	1.105	0.226	.638	0.002
Time ∗ ExpGroup	0.001	1	0.001	2.689e -4	.987	2.303e -6
Residuals	131.895	27	4.885			

Cases	Sum of Squares	df	Mean Square	F	*p*	η²
ExpGroup	72.854	1	72.854	5.395	.028	0.128
Residuals	364.629	27	13.505			

Post Hoc Comparisons - ExpGroup

			95% CI for Mean Difference				
		Mean Difference	**Lower**	**Upper**	**SE**	**t**	***p* bonf**
Collab	FFI+Collab	-2.243	-4.224	-0.262	0.966	-2.323	.028

Mechanics

Cases	Sum of Squares	df	Mean Square	F	*p*	η²
Time	17.524	1	17.524	2.110	.158	0.022
Time ∗ ExpGroup	17.524	1	17.524	2.110	.158	0.022
Residuals	224.200	27	8.304			

Cases	Sum of Squares	df	Mean Square	F	*p*	η²
ExpGroup	3.974	1	3.974	0.204	.655	0.005
Residuals	526.095	27	19.485			

Discourse

Cases	Sum of Squares	df	Mean Square	F	*p*	η²
Time	0.512	1	0.512	0.418	.523	0.005
Time ∗ ExpGroup	2.995	1	2.995	2.445	.130	0.030
Residuals	33.074	27	1.225			

Cases	Sum of Squares	df	Mean Square	F	*p*	η²
ExpGroup	0.838	1	0.838	0.367	.550	0.008
Residuals	61.645	27	2.283			

Repeated measures ANOVA on LRE outcome

Correct

Cases	Sum of Squares	df	Mean Square	F	*p*	η²
Time	10.522	1	10.522	0.529	.473	0.004
Time ✻ ExpGroup	42.798	1	42.798	2.154	.154	0.017
Residuals	536.581	27	19.873			

Cases	Sum of Squares	df	Mean Square	F	*p*	η²
ExpGroup	22.200	1	22.200	0.307	.584	0.009
Residuals	1953.524	27	72.353			

Incorrect

Cases	Sum of Squares	df	Mean Square	F	*p*	η²
Time	4.268	1	4.268	1.704	.203	0.015
Time ✻ ExpGroup	7.992	1	7.992	3.191	.085	0.029
Residuals	67.629	27	2.505			

Cases	Sum of Squares	df	Mean Square	F	*p*	η²
ExpGroup	16.626	1	16.626	2.510	.125	0.060
Residuals	178.857	27	6.624			

Addressed

Cases	Sum of Squares	df	Mean Square	F	*p*	η²
Time	6.074	1	6.074	4.609	.041	0.029
Time ✻ ExpGroup	0.695	1	0.695	0.527	.474	0.003
Residuals	35.581	27	1.318			

Cases	Sum of Squares	df	Mean Square	F	*p*	η²
ExpGroup	15.181	1	15.181	2.749	.109	0.073
Residuals	149.095	27	5.522			

Post-hoc comparisons - Time

		95% CI for Mean Difference				
	Mean Difference	Lower	Upper	SE	t	p_{bonf}
Day.1 Day.2	0.648	0.029	1.267	0.302	2.147	.041

Ignored

Cases	Sum of Squares	df	Mean Square	F	*p*	η²
Time	0.089	1	0.089	0.047	.831	3.121e -4
Time ✻ ExpGroup	1.124	1	1.124	0.586	.451	0.004
Residuals	51.807	27	1.919			

Cases	Sum of Squares	df	Mean Square	F	*p*	η²
ExpGroup	0.152	1	0.152	0.018	.896	5.299e -4
Residuals	233.331	27	8.642			

Repeated measures ANOVA on LRE engagement

Elaborate

Cases	Sum of Squares	df	Mean Square	F	*p*	η²
Time	38.975	1	38.975	3.723	.064	0.022
Time ✻ ExpGroup	29.458	1	29.458	2.814	.105	0.017
Residuals	282.645	27	10.468			

Cases	Sum of Squares	df	Mean Square	F	*p*	η²
ExpGroup	226.512	1	226.512	5.191	.031	0.129
Residuals	1178.074	27	43.632			

Post Hoc Comparisons - ExpGroup

		95% CI for Mean Difference				
	Mean Difference	Lower	Upper	SE	t	p_{bonf}
Collab FFI+Collab	-3.955	-7.516	-0.393	1.736	-2.278	.031

Simple

Cases	Sum of Squares	df	Mean Square	F	*p*	η²
Time	4.268	1	4.268	1.704	.203	0.015
Time ✻ ExpGroup	7.992	1	7.992	3.191	.085	0.029
Residuals	67.629	27	2.505			

Cases	Sum of Squares	df	Mean Square	F	*p*	η²
ExpGroup	16.626	1	16.626	2.510	.125	0.060
Residuals	178.857	27	6.624			

Repeated measures ANOVA on length of PKL and L2 turns

PKL

Cases	Sum of Squares	df	Mean Square	F	*p*	η²
Time	2.635	1	2.635	0.427	.519	0.005
Time ✻ expgroup	17.542	1	17.542	2.839	.104	0.035
Residuals	166.814	27	6.178			

Cases	Sum of Squares	df	Mean Square	F	*p*	η²
expgroup	72.564	1	72.564	8.233	.008	0.146
Residuals	237.963	27	8.813			

Post Hoc Comparisons - expgroup

		95% CI for Mean Difference				
	Mean Difference	Lower	Upper	SE	t	p_{bonf}
Collab FFI+Collab	-2.238	-3.839	-0.638	0.780	-2.869	.008

L2

Cases	Sum of Squares	df	Mean Square	F	*p*	η²
Time	0.318	1	0.318	0.048	.829	6.082e -4
Time ✻ expgroup	3.068	1	3.068	0.460	.503	0.006
Residuals	179.909	27	6.663			

Cases	Sum of Squares	df	Mean Square	F	*p*	η²
expgroup	3.319	1	3.319	0.266	.610	0.006
Residuals	337.029	27	12.483			

B3 Inferential statistics for RQ1d

Two-way ANOVA on 3S

Cases	Sum of Squares	df	Mean Square	F	*p*	η^2
Target	16.478	1	16.478	5.831	.019	0.086
ExpGroup	7.000	1	7.000	2.477	.121	0.036
Target ✻ ExpGroup	16.478	1	16.478	5.831	.019	0.086
Residuals	152.590	54	2.826			

Post Hoc Comparisons - Target

		Mean Difference	95% CI for Mean Difference Lower	Upper	SE	t	p_{bonf}
3S	POSS	1.067	0.181	1.952	0.442	2.415	.019

Post Hoc Comparisons - Target ✻ ExpGroup

		Mean Difference	95% CI for Mean Difference Lower	Upper	SE	t	p_{bonf}
3S, Collab	POSS, Collab	1.776e -15	-1.684	1.684	0.635	2.796e -15	1.000
	3S, FFI+Collab	-1.762	-3.418	-0.106	0.625	-2.821	.040
	POSS, FFI+Collab	0.371	-1.285	2.027	0.625	0.595	1.000
POSS, Collab	3S, FFI+Collab	-1.762	-3.418	-0.106	0.625	-2.821	.040
	POSS, FFI+Collab	0.371	-1.285	2.027	0.625	0.595	1.000
3S, FFI+Collab	POSS, FFI+Collab	2.133	0.506	3.760	0.614	3.476	.006

Two-way ANOVA on POSS

Cases	Sum of Squares	df	Mean Square	F	*p*	η^2
Target	9.716	1	9.716	7.630	.008	0.122
ExpGroup	0.005	1	0.005	0.004	.949	6.620e -5
Target ✻ ExpGroup	0.888	1	0.888	0.697	.407	0.011
Residuals	68.762	54	1.273			

Post Hoc Comparisons - Target

		95% CI for Mean Difference				
	Mean Difference	Lower	Upper	SE	t	p_{bonf}
3S POSS	-0.819	-1.414	-0.225	0.297	-2.762	.008

Paired samples t-test for COLLAB

Measure 1		Measure 2	t	df	*p*	Cohen's *d*
@3s_TARGET	-	POSS_TARGET	-1.194	13	0.254	-0.319

Paired Samples T-Test for FFI+COLLAB

Measure 1		Measure 2	t	df	*p*	Cohen's *d*
@3s_TARGET	-	POSS_TARGET	1.586	14	0.135	0.410

APPENDIX C

C1 Statistical analyses for RQ2b

Clause type

Preclauses

Within Subjects Effects

Cases	Sum of Squares	df	Mean Square	F	*p*	η²
Time	475.077	1	475.077	7.743	0.007	0.039
Time ✻ EXPGROUP	253.703	2	126.852	2.067	0.137	0.021
Residuals	3190.606	52	61.358			

Between Subjects Effects

Cases	Sum of Squares	df	Mean Square	F	*p*	η²
EXPGROUP	549.736	2	274.868	1.840	0.169	0.045
Residuals	7766.918	52	149.364			

Post Hoc Comparisons - Time

			95% CI for Mean Difference					
		Mean Difference	Lower	Upper	SE	t	Cohen's *d*	*p* bonf
Day.1	Day.2	-4.317	-7.431	-1.204	1.552	-2.783	-0.375	0.007

Protoclauses

Within Subjects Effects

Cases	Sum of Squares	df	Mean Square	F	*p*	η²
Time	5808.780	1	5808.780	13.590	< .001	0.068
Time ✻ EXPGROUP	342.384	2	171.192	0.401	0.672	0.004
Residuals	22226.977	52	427.442			

Between Subjects Effects

Cases	Sum of Squares	df	Mean Square	F	*p*	η²
EXPGROUP	5180.653	2	2590.327	2.570	0.086	0.060
Residuals	52410.707	52	1007.898			

Post Hoc Comparisons - Time

			95% CI for Mean Difference					
		Mean Difference	Lower	Upper	SE	t	Cohen's *d*	*p* bonf
Day.1	Day.2	-15.097	-23.314	-6.879	4.095	-3.686	-0.497	< .001

Paraclause

Within Subjects Effects

Cases	Sum of Squares	df	Mean Square	F	*p*	η²
Time	57.059	1	57.059	2.757	0.103	0.022
Time ✻ EXPGROUP	4.246	2	2.123	0.103	0.903	0.002
Residuals	1076.371	52	20.699			

Between Subjects Effects

Cases	Sum of Squares	df	Mean Square	F	*p*	η²
EXPGROUP	45.634	2	22.817	0.837	0.439	0.018
Residuals	1418.124	52	27.272			

Full-fledged clauses

Within Subjects Effects

Cases	Sum of Squares	df	Mean Square	F	*p*	η²
Time	11144.052	1	11144.052	24.326	< .001	0.096
Time ✻ EXPGROUP	1061.099	2	530.550	1.158	0.322	0.009
Residuals	23822.123	52	458.118			

Between Subjects Effects

Cases	Sum of Squares	df	Mean Square	F	*p*	η²
EXPGROUP	6959.519	2	3479.759	2.459	0.095	0.060
Residuals	73576.757	52	1414.938			

Post Hoc Comparisons - Time

			95% CI for Mean Difference					
		Mean Difference	Lower	Upper	SE	t	Cohen's *d*	p_{bonf}
Day.1	Day.2	20.910	12.403	29.418	4.240	4.932	0.665	< .001

Complexity and accuracy

Clauses/T-unit

Within Subjects Effects

Cases	Sum of Squares	df	Mean Square	F	*p*	η²
Time	0.024	1	0.024	2.157	0.148	0.020
Time ✻ EXPGROUP	0.005	2	0.003	0.237	0.790	0.004
Residuals	0.584	52	0.011			

Between Subjects Effects

Cases	Sum of Squares	df	Mean Square	F	*p*	η²
EXPGROUP	0.024	2	0.012	1.144	0.327	0.020
Residuals	0.551	52	0.011			

Coord/T-unit

Within Subjects Effects

Cases	Sum of Squares	df	Mean Square	F	*p*	η²
Time	0.358	1	0.358	4.367	0.042	0.037
Time ✻ EXPGROUP	0.028	2	0.014	0.173	0.842	0.003
Residuals	4.259	52	0.082			

Between Subjects Effects

Cases	Sum of Squares	df	Mean Square	F	*p*	η²
EXPGROUP	0.570	2	0.285	3.249	0.047	0.058
Residuals	4.561	52	0.088			

Post Hoc Comparisons - Time

			95% CI for Mean Difference					
		Mean Difference	Lower	Upper	SE	t	Cohen's *d*	p_{bonf}
Day.1	Day.2	0.118	0.005	0.232	0.057	2.090	0.282	0.042

Post Hoc Comparisons - EXPGROUP

		Mean Difference	95% CI for Mean Difference Lower	Upper	SE	t	Cohen's *d*	*p* bonf
Comp	Collab	-0.113	-0.284	0.059	0.069	-1.623	-0.219	0.332
	FFI+Collab	-0.164	-0.332	0.004	0.068	-2.416	-0.326	0.058
Collab	FFI+Collab	-0.051	-0.244	0.141	0.078	-0.661	-0.089	1.000

MLC

Within Subjects Effects

Cases	Sum of Squares	df	Mean Square	F	p	η²
Time	0.278	1	0.278	0.122	0.728	9.785e -4
Time ✻ EXPGROUP	3.820	2	1.910	0.843	0.436	0.013
Residuals	117.840	52	2.266			

Between Subjects Effects

Cases	Sum of Squares	df	Mean Square	F	*p*	η²
EXPGROUP	7.963	2	3.982	1.347	0.269	0.028
Residuals	153.762	52	2.957			

GI

Within Subjects Effects

Cases	Sum of Squares	df	Mean Square	F	*p*	η²
Time	0.612	1	0.612	2.263	0.139	0.005
Time ✻ EXPGROUP	0.777	2	0.388	1.437	0.247	0.006
Residuals	14.054	52	0.270			

Between Subjects Effects

Cases	Sum of Squares	df	Mean Square	F	*p*	η²
EXPGROUP	6.809	2	3.405	1.598	0.212	0.051
Residuals	110.758	52	2.130			

GrammErr100

Within Subjects Effects

Cases	Sum of Squares	df	Mean Square	F	*p*	η²
Time	200.571	1	200.571	5.920	0.018	0.032
Time ✻ EXPGROUP	138.917	2	69.458	2.050	0.139	0.022
Residuals	1761.681	52	33.878			

Between Subjects Effects

Cases	Sum of Squares	df	Mean Square	F	*p*	η²
EXPGROUP	25.028	2	12.514	0.155	0.857	0.004
Residuals	4203.034	52	80.828			

Post Hoc Comparisons - Time

			95% CI for Mean Difference					
		Mean Difference	Lower	Upper	SE	t	Cohen's *d*	*p* bonf
Day.1	Day.2	2.805	0.492	5.119	1.153	2.433	0.328	0.018

GramErrFREEC

Within Subjects Effects

Cases	Sum of Squares	df	Mean Square	F	*p*	η²
Time	1398.453	1	1398.453	3.471	0.068	0.020
Time ✻ EXPGROUP	1712.552	2	856.276	2.126	0.130	0.024
Residuals	20948.354	52	402.853			

Between Subjects Effects

Cases	Sum of Squares	df	Mean Square	F	*p*	η²
EXPGROUP	1286.707	2	643.353	0.744	0.480	0.018
Residuals	44979.999	52	865.000			

SpellErr100

Within Subjects Effects

Cases	Sum of Squares	df	Mean Square	F	*p*	η²
Time	21.001	1	21.001	0.796	0.376	0.003
Time ✻ EXPGROUP	98.411	2	49.206	1.864	0.165	0.014
Residuals	1372.370	52	26.392			

Between Subjects Effects

Cases	Sum of Squares	df	Mean Square	F	*p*	η²
EXPGROUP	468.423	2	234.212	2.295	0.111	0.064
Residuals	5306.743	52	102.053			

Borrow100

Within Subjects Effects

Cases	Sum of Squares	df	Mean Square	F	*p*	η²
Time	8.090e -4	1	8.090e -4	3.834e -5	0.995	3.270e -7
Time ✻ EXPGROUP	11.469	2	5.735	0.272	0.763	0.005
Residuals	1097.267	52	21.101			

Between Subjects Effects

Cases	Sum of Squares	df	Mean Square	F	*p*	η²
EXPGROUP	59.100	2	29.550	1.177	0.316	0.024
Residuals	1306.036	52	25.116			

3SAcc

ANOVA - @3Ssupp

Cases	Sum of Squares	df	Mean Square	F	*p*	η²
EXPGROUP	3768.803	2	1884.402	2.229	0.119	0.087
Residuals	39738.403	47	845.498			

POSSAcc

ANOVA - POSSsupp

Homogeneity Correction	Cases	Sum of Squares	df	Mean Square	F	*p*	η²
None	EXPGROUP	5801.942	2.000	2900.971	3.819	0.029	0.133
	Residuals	37985.060	50.000	759.701			
Brown-Forsythe	EXPGROUP	5801.942	2.000	2900.971	2.769	0.086	0.133
	Residuals	37985.060	21.135	1797.233			

IU (Content accuracy)

Within Subjects Effects

Cases	Sum of Squares	df	Mean Square	F	*p*	η²
Time	0.837	1	0.837	0.230	0.633	7.813e -4
Time ✻ EXPGROUP	6.428	2	3.214	0.885	0.419	0.006
Residuals	188.790	52	3.631			

Between Subjects Effects

Cases	Sum of Squares	df	Mean Square	F	*p*	η²
EXPGROUP	79.506	2	39.753	2.599	0.084	0.074
Residuals	795.458	52	15.297			

Rubric measures

Adq

Within Subjects Effects

Cases	Sum of Squares	df	Mean Square	F	*p*	η²
Time	0.198	1	0.198	0.711	0.403	0.003
Time ✻ EXPGROUP	0.566	2	0.283	1.016	0.369	0.008
Residuals	14.489	52	0.279			

Between Subjects Effects

Cases	Sum of Squares	df	Mean Square	F	*p*	η²
EXPGROUP	3.614	2	1.807	1.729	0.188	0.049
Residuals	54.349	52	1.045			

Coher

Within Subjects Effects

Cases	Sum of Squares	df	Mean Square	F	*p*	η^2
Time	0.022	1	0.022	0.074	0.787	3.422e -4
Time ✻ EXPGROUP	0.259	2	0.130	0.445	0.644	0.004
Residuals	15.159	52	0.292			

Between Subjects Effects

Cases	Sum of Squares	df	Mean Square	F	*p*	η^2
EXPGROUP	4.709	2	2.355	2.865	0.066	0.075
Residuals	42.745	52	0.822			

Cohes

Within Subjects Effects

Cases	Sum of Squares	df	Mean Square	F	*p*	η^2
Time	0.554	1	0.554	2.197	0.144	0.009
Time ✻ EXPGROUP	0.547	2	0.273	1.083	0.346	0.009
Residuals	13.126	52	0.252			

Between Subjects Effects

Cases	Sum of Squares	df	Mean Square	F	*p*	η^2
EXPGROUP	0.687	2	0.343	0.394	0.676	0.011
Residuals	45.277	52	0.871			

Acc

Within Subjects Effects

Cases	Sum of Squares	df	Mean Square	F	*p*	η^2
Time	0.524	1	0.524	3.236	0.078	0.012
Time ✻ EXPGROUP	0.002	2	0.001	0.006	0.994	4.835e -5
Residuals	8.416	52	0.162			

Between Subjects Effects

Cases	Sum of Squares	df	Mean Square	F	*p*	η^2
EXPGROUP	0.132	2	0.066	0.102	0.903	0.003
Residuals	33.632	52	0.647			

Mech

Within Subjects Effects

Cases	Sum of Squares	df	Mean Square	F	*p*	η^2
Time	1.379	1	1.379	3.924	0.053	0.018
Time ✻ EXPGROUP	0.692	2	0.346	0.984	0.381	0.009
Residuals	18.272	52	0.351			

Between Subjects Effects

Cases	Sum of Squares	df	Mean Square	F	*p*	η^2
EXPGROUP	10.034	2	5.017	5.624	0.006	0.131
Residuals	46.384	52	0.892			

Post Hoc Comparisons - EXPGROUP

		Mean Difference	SE	t	Cohen's *d*	*p* bonf
Comp	Collab	-0.486	0.221	-2.196	-0.296	0.098
	FFI+Collab	0.342	0.217	1.581	0.213	0.360
Collab	FFI+Collab	0.829	0.248	3.339	0.450	0.005

Lex

Within Subjects Effects

Cases	Sum of Squares	df	Mean Square	F	*p*	η²
Time	0.084	1	0.084	0.326	0.571	0.001
Time ✻ EXPGROUP	0.071	2	0.035	0.138	0.872	0.001
Residuals	13.347	52	0.257			

Between Subjects Effects

Cases	Sum of Squares	df	Mean Square	F	*p*	η²
EXPGROUP	4.901	2	2.450	2.737	0.074	0.075
Residuals	46.554	52	0.895			

C2 Statistical analyses for RQ2c

Repeated Measures ANOVA for clause type

PreCRatio

Within Subjects Effects

Cases	Sum of Squares	df	Mean Square	F	*p*	η^2
Time	3437.867 [a]	2 [a]	1718.934 [a]	3.613 [a]	0.029 [a]	0.018
Time ✻ expgroup	2496.856 [a]	4 [a]	624.214 [a]	1.312 [a]	0.268 [a]	0.013
Residuals	71363.666	150	475.758			

Between Subjects Effects

Cases	Sum of Squares	df	Mean Square	F	*p*	η^2
expgroup	5700.681	2	2850.340	2.070	0.133	0.031
Residuals	103250.161	75	1376.669			

Post Hoc Comparisons - Time

		Mean Difference	95% CI for Mean Difference		SE	t	Cohen's *d*	p_{bonf}
			Lower	Upper				
T1	T2	9.386	0.918	17.854	3.498	2.684	0.304	0.024
	T3	4.217	-4.251	12.685	3.498	1.206	0.137	0.690
T2	T3	-5.169	-13.638	3.299	3.498	-1.478	-0.167	0.425

ProtoCRatio

Within Subjects Effects

Cases	Sum of Squares	df	Mean Square	F	*p*	η^2
Time	547.881 [a]	2 [a]	273.940 [a]	0.435 [a]	0.648 [a]	0.003
Time ✻ expgroup	2984.429 [a]	4 [a]	746.107 [a]	1.186 [a]	0.319 [a]	0.014
Residuals	94378.611	150	629.191			

Between Subjects Effects

Cases	Sum of Squares	df	Mean Square	F	*p*	η^2
expgroup	1423.719	2	711.860	0.451	0.639	0.007
Residuals	118370.487	75	1578.273			

ParaCRatio

Within Subjects Effects

Cases	Sum of Squares	df	Mean Square	F	*p*	η^2
Time	180.091 [a]	2 [a]	90.046 [a]	1.359 [a]	0.260 [a]	0.011
Time ✻ expgroup	205.586 [a]	4 [a]	51.396 [a]	0.776 [a]	0.543 [a]	0.013
Residuals	9940.057	150	66.267			

Between Subjects Effects

Cases	Sum of Squares	df	Mean Square	F	*p*	η^2
expgroup	136.524	2	68.262	0.886	0.417	0.008
Residuals	5781.580	75	77.088			

FullCRatio

Within Subjects Effects

Cases	Sum of Squares	df	Mean Square	F	*p*	η^2
Time	665.997	2	332.999	1.252	0.289	0.003
Time ✻ expgroup	4703.999	4	1176.000	4.421	0.002	0.019
Residuals	39900.895	150	266.006			

Between Subjects Effects

Cases	Sum of Squares	df	Mean Square	F	*p*	η^2
expgroup	5000.037	2	2500.019	0.932	0.398	0.020
Residuals	201203.444	75	2682.713			

Post Hoc Comparisons - expgroup ✻ Time

		Mean Difference	95% CI for Mean Difference		SE	t	p bonf
			Lower	Upper			
Comp, T1	Collab, T1	-17.946	-48.352	12.459	9.259	-1.938	1.000
	FFI+Collab, T1	6.752	-22.827	36.332	9.007	0.750	1.000
	Comp, T2	-3.481	-18.509	11.548	4.613	-0.755	1.000
	Collab, T2	-13.061	-43.467	17.344	9.259	-1.411	1.000
	FFI+Collab, T2	-6.217	-35.797	23.363	9.007	-0.690	1.000
	Comp, T3	-6.410	-21.439	8.618	4.613	-1.390	1.000
	Collab, T3	-6.954	-37.359	23.452	9.259	-0.751	1.000
	FFI+Collab, T3	-7.522	-37.102	22.058	9.007	-0.835	1.000
Collab, T1	FFI+Collab, T1	24.699	-4.881	54.279	9.007	2.742	0.258
	Comp, T2	14.466	-15.940	44.871	9.259	1.562	1.000
	Collab, T2	4.885	-10.143	19.913	4.613	1.059	1.000
	FFI+Collab, T2	11.730	-17.850	41.310	9.007	1.302	1.000
	Comp, T3	11.536	-18.869	41.942	9.259	1.246	1.000
	Collab, T3	10.992	-4.036	26.021	4.613	2.383	0.663
	FFI+Collab, T3	10.424	-19.156	40.004	9.007	1.157	1.000
FFI+Collab, T1	Comp, T2	-10.233	-39.813	19.347	9.007	-1.136	1.000
	Collab, T2	-19.814	-49.394	9.766	9.007	-2.200	1.000
	FFI+Collab, T2	-12.969	-27.170	1.231	4.359	-2.975	0.123
	Comp, T3	-13.163	-42.743	16.417	9.007	-1.461	1.000
	Collab, T3	-13.706	-43.286	15.874	9.007	-1.522	1.000
	FFI+Collab, T3	-14.275	-28.475	-0.074	4.359	-3.275	.047 *
Comp, T2	Collab, T2	-9.581	-39.986	20.825	9.259	-1.035	1.000
	FFI+Collab, T2	-2.736	-32.316	26.844	9.007	-0.304	1.000
	Comp, T3	-2.930	-17.958	12.099	4.613	-0.635	1.000
	Collab, T3	-3.473	-33.879	26.932	9.259	-0.375	1.000
	FFI+Collab, T3	-4.042	-33.622	25.538	9.007	-0.449	1.000
Collab, T2	FFI+Collab, T2	6.845	-22.735	36.425	9.007	0.760	1.000

Post Hoc Comparisons - expgroup ✻ Time

		Mean Difference	95% CI for Mean Difference		SE	t	p_{bonf}
			Lower	Upper			
	Comp, T3	6.651	-23.754	37.057	9.259	0.718	1.000
	Collab, T3	6.108	-8.921	21.136	4.613	1.324	1.000
	FFI+Collab, T3	5.539	-24.041	35.119	9.007	0.615	1.000
FFI+Collab, T2	Comp, T3	-0.194	-29.773	29.386	9.007	-0.021	1.000
	Collab, T3	-0.737	-30.317	28.843	9.007	-0.082	1.000
	FFI+Collab, T3	-1.306	-15.506	12.895	4.359	-0.300	1.000
Comp, T3	Collab, T3	-0.544	-30.949	29.862	9.259	-0.059	1.000
	FFI+Collab, T3	-1.112	-30.692	28.468	9.007	-0.123	1.000
Collab, T3	FFI+Collab, T3	-0.568	-30.148	29.012	9.007	-0.063	1.000

* p < .05

Repeated Measures ANOVA for complexity and general accuracy

Clauses/T-unit

Within Subjects Effects

Cases	Sum of Squares	df	Mean Square	F	*p*	η²
Time	0.042	2	0.021	1.824	0.165	0.009
Time ✻ expgroup	0.072	4	0.018	1.576	0.184	0.016
Residuals	1.623	142	0.011			

Between Subjects Effects

Cases	Sum of Squares	df	Mean Square	F	*p*	η²
expgroup	0.056	2	0.028	0.731	0.485	0.012
Residuals	2.704	71	0.038			

Within Subjects Effects

Cases	Sum of Squares	df	Mean Square	F	*p*	η²
Time	0.062	2	0.031	0.462	0.631	0.003
Time ✻ expgroup	0.163	4	0.041	0.606	0.659	0.008
Residuals	9.570	142	0.067			

Coord/T-unnit

Between Subjects Effects

Cases	Sum of Squares	df	Mean Square	F	*p*	η²
expgroup	0.051	2	0.026	0.174	0.841	0.003
Residuals	10.458	71	0.147			

Within Subjects Effects

Cases	Sum of Squares	df	Mean Square	F	*p*	η²
Time	19.744 [a]	2 [a]	9.872 [a]	2.580 [a]	0.079 [a]	0.020
Time ✻ expgroup	10.380 [a]	4 [a]	2.595 [a]	0.678 [a]	0.608 [a]	0.010
Residuals	543.343	142	3.826			

MLC

Within Subjects Effects

Cases	Sum of Squares	df	Mean Square	F	*p*	η²
Time	19.744 [a]	2 [a]	9.872 [a]	2.580 [a]	0.079 [a]	0.020
Time ✻ expgroup	10.380 [a]	4 [a]	2.595 [a]	0.678 [a]	0.608 [a]	0.010
Residuals	543.343	142	3.826			

Between Subjects Effects

Cases	Sum of Squares	df	Mean Square	F	*p*	η²
expgroup	3.919	2	1.959	0.333	0.718	0.004
Residuals	417.343	71	5.878			

GI

Within Subjects Effects

Cases	Sum of Squares	df	Mean Square	F	*p*	η²
Time	0.291	2	0.145	0.576	0.563	0.001
Time ✻ expgroup	1.298	4	0.325	1.288	0.277	0.006
Residuals	37.808	150	0.252			

Between Subjects Effects

Cases	Sum of Squares	df	Mean Square	F	*p*	η²
expgroup	33.559	2	16.779	9.828	< .001	0.167
Residuals	128.051	75	1.707			

Post Hoc Comparisons - expgroup

		Mean Difference	95% CI for Mean Difference Lower	95% CI for Mean Difference Upper	SE	t	Cohen's *d*	*p* bonf
Comp	Collab	-0.404	-0.927	0.119	0.213	-1.893	-0.214	0.186
	FFI+Collab	0.512	0.003	1.020	0.208	2.466	0.279	0.048
Collab	FFI+Collab	0.916	0.408	1.424	0.208	4.412	0.500	< .001

GramErr100

Within Subjects Effects

Cases	Sum of Squares	df	Mean Square	F	*p*	η²
Time	172.495	2	86.247	1.606	0.204	0.009
Time ✻ expgroup	184.552	4	46.138	0.859	0.490	0.010
Residuals	8055.203	150	53.701			

Between Subjects Effects

Cases	Sum of Squares	df	Mean Square	F	*p*	η²
expgroup	341.341	2	170.670	1.282	0.284	0.018
Residuals	9985.973	75	133.146			

GramErrFREEClauses

Within Subjects Effects

Cases	Sum of Squares	df	Mean Square	F	*p*	η²
Time	551.248	2	275.624	1.055	0.351	0.005
Time ✻ expgroup	1012.385	4	253.096	0.969	0.427	0.009
Residuals	37102.302	142	261.284			

Between Subjects Effects

Cases	Sum of Squares	df	Mean Square	F	*p*	η²
expgroup	1895.723	2	947.862	0.916	0.405	0.017
Residuals	73508.299	71	1035.328			

SpellErr100

Within Subjects Effects

Cases	Sum of Squares		df	Mean Square		F	p	η²
Time	104.092	a	2 a	52.046	a	0.746 a	0.476 a	0.005
Time ✻ expgroup	231.361	a	4 a	57.840	a	0.829 a	0.509 a	0.010
Residuals	10468.251		150	69.788				

Between Subjects Effects

Cases	Sum of Squares	df	Mean Square	F	p	η²
expgroup	168.618	2	84.309	0.554	0.577	0.008
Residuals	11420.682	75	152.276			

Borrow100

Within Subjects Effects

Cases	Sum of Squares		df	Mean Square		F	p	η²
Time	458.112	a	2 a	229.056	a	3.191 a	0.044 a	0.018
Time ✻ expgroup	97.381	a	4 a	24.345	a	0.339 a	0.851 a	0.004
Residuals	10766.012		150	71.773				

Between Subjects Effects

Cases	Sum of Squares	df	Mean Square	F	p	η²
expgroup	664.014	2	332.007	1.851	0.164	0.026
Residuals	13454.279	75	179.390			

Post Hoc Comparisons - Time

			95% CI for Mean Difference					
		Mean Difference	Lower	Upper	SE	t	Cohen's *d*	p_{bonf}
T1	T2	-0.555	-3.844	2.734	1.359	-0.409	-0.046	1.000
	T3	-3.211	-6.500	0.078	1.359	-2.363	-0.268	0.058
T2	T3	-2.656	-5.945	0.633	1.359	-1.955	-0.221	0.157

One-way ANOVA for 3SAcc

T1

ANOVA - @3SAcc.1

Cases	Sum of Squares	df	Mean Square	F	*p*	η^2
expgroup	4697.178	2	2348.589	2.805	0.068	0.083
Residuals	51904.349	62	837.167			

T2

ANOVA - @3SAcc.2

Cases	Sum of Squares	df	Mean Square	F	*p*	η^2
expgroup	2288.413	2	1144.206	2.034	0.139	0.062
Residuals	34876.160	62	562.519			

T3

ANOVA - @3SAcc.3

Cases	Sum of Squares	df	Mean Square	F	*p*	η^2
expgroup	2370.044	2	1185.022	1.449	0.242	0.043
Residuals	52345.126	64	817.893			

One-way ANOVA for POSSAcc

T1

ANOVA - @POSSAcc.1

Cases	Sum of Squares	df	Mean Square	F	*p*	η^2
expgroup	1124.152	2	562.076	0.330	0.721	0.021
Residuals	52783.120	31	1702.681			

T2

ANOVA - @POSSAcc.2

Cases	Sum of Squares	df	Mean Square	F	*p*	η²
expgroup	3368.599	2	1684.299	0.952	0.396	0.050
Residuals	63711.031	36	1769.751			

T3

ANOVA - @POSSAcc.3

Cases	Sum of Squares	df	Mean Square	F	*p*	η²
expgroup	15167.411	2	7583.705	4.430	0.023	0.270
Residuals	41082.589	24	1711.775			

Post Hoc Comparisons - expgroup

		Mean Difference	95% CI for Mean Difference		SE	t	Cohen's *d*	*p* bonf
			Lower	Upper				
Comp	Collab	3.571	-45.568	52.711	19.677	0.182	0.083	1.000
	FFI+Collab	-49.554	-103.028	3.921	21.413	-2.314	-1.163	0.089
Collab	FFI+Collab	-53.125	-100.285	-5.965	18.884	-2.813	-1.377	0.029

Repeated Measures ANOVA for 3S vs POSS

At T1

Within Subjects Effects

Cases	Sum of Squares	df	Mean Square	F	*p*	η²
Target	155.690	1	155.690	28.863	< .001	0.210
Target ✻ expgroup	41.150	2	20.575	3.814	0.033	0.056
Residuals	167.218	31	5.394			

Between Subjects Effects

Cases	Sum of Squares	df	Mean Square	F	*p*	η^2
expgroup	44.455	2	22.228	2.077	0.142	0.060
Residuals	331.795	31	10.703			

Post Hoc Comparisons - expgroup ✻ Target

		Mean Difference	95% CI for Mean Difference		SE	t	p_{bonf}	
			Lower	Upper				
Comp, X3S	Collab, X3S	-2.917	-6.753	0.920	1.251	-2.331	0.257	
	FFI+Collab, X3S	0.615	-3.157	4.388	1.230	0.500	1.000	
	Comp, POSS	2.333	-1.148	5.815	1.095	2.131	0.411	
	Collab, POSS	2.250	-1.586	6.086	1.251	1.799	0.593	
	FFI+Collab, POSS	2.308	-1.465	6.080	1.230	1.876	0.593	
Collab, X3S	FFI+Collab, X3S	3.532	0.049	7.015	1.136	3.110	0.035	*
	Comp, POSS	5.250	1.414	9.086	1.251	4.197	0.001	**
	Collab, POSS	5.167	2.151	8.182	0.948	5.449	< .001	***
	FFI+Collab, POSS	5.224	1.742	8.707	1.136	4.600	< .001	***
FFI+Collab, X3S	Comp, POSS	1.718	-2.055	5.490	1.230	1.396	0.934	
	Collab, POSS	1.635	-1.848	5.117	1.136	1.439	0.934	
	FFI+Collab, POSS	1.692	-1.205	4.589	0.911	1.858	0.593	
Comp, POSS	Collab, POSS	-0.083	-3.920	3.753	1.251	-0.067	1.000	
	FFI+Collab, POSS	-0.026	-3.798	3.747	1.230	-0.021	1.000	
Collab, POSS	FFI+Collab, POSS	0.058	-3.425	3.540	1.136	0.051	1.000	

* p < .05, ** p < .01, *** p < .001

At T2

Within Subjects Effects

Cases	Sum of Squares	df	Mean Square	F	*p*	η²
Target	228.359	1	228.359	25.720	< .001	0.241
Target ✻ expgroup	11.001	2	5.500	0.620	0.544	0.012
Residuals	310.749	35	8.879			

Between Subjects Effects

Cases	Sum of Squares	df	Mean Square	F	*p*	η²
expgroup	19.041	2	9.520	0.878	0.425	0.020
Residuals	379.446	35	10.841			

Post Hoc Comparisons - Target

		95% CI for Mean Difference				
	Mean Difference	Lower	Upper	SE	t	*p* bonf
X3S POSS	3.496	2.097	4.896	0.689	5.072	< .001 ***

*** p < .001

At T3

Within Subjects Effects

Cases	Sum of Squares	df	Mean Square	F	*p*	η²
Target	182.758	1	182.758	50.194	< .001	0.387
Target ✻ expgroup	5.436	2	2.718	0.746	0.484	0.012
Residuals	94.668	26	3.641			

Between Subjects Effects

Cases	Sum of Squares	df	Mean Square	F	*p*	η²
expgroup	13.781	2	6.890	1.019	0.375	0.029
Residuals	175.840	26	6.763			

Repeated Measures ANOVA for 3S across TIME

Within Subjects Effects

Cases	Sum of Squares	df	Mean Square	F	*p*	η²
Time	8.813	2	4.407	0.012	0.988	7.956e -5
Time ✻ expgroup	3141.420	4	785.355	2.181	0.078	0.028
Residuals	31681.698	88	360.019			

Between Subjects Effects

Cases	Sum of Squares	df	Mean Square	F	*p*	η²
expgroup	3853.169	2	1926.584	1.176	0.318	0.035
Residuals	72088.825	44	1638.382			

Repeated measures ANOVA for rubric measures

Adq

Within Subjects Effects

Cases	Sum of Squares	df	Mean Square	F	*p*	η^2
Time	2.276	2	1.138	4.349	0.015	0.018
Time ✻ expgroup	1.003	4	0.251	0.959	0.432	0.008
Residuals	39.245	150	0.262			

Between Subjects Effects

Cases	Sum of Squares	df	Mean Square	F	*p*	η^2
expgroup	9.281	2	4.640	4.760	0.011	0.074
Residuals	73.112	75	0.975			

Post Hoc Comparisons - Time

			95% CI for Mean Difference					
		Mean Difference	Lower	Upper	SE	t	Cohen's *d*	p_{bonf}
T1	T2	-0.236	-0.435	-0.038	0.082	-2.880	-0.326	0.014
	T3	-0.163	-0.362	0.035	0.082	-1.991	-0.225	0.145
T2	T3	0.073	-0.126	0.271	0.082	0.888	0.101	1.000

Post Hoc Comparisons - expgroup

			95% CI for Mean Difference					
		Mean Difference	Lower	Upper	SE	t	Cohen's *d*	p_{bonf}
Comp	Collab	-0.267	-0.661	0.128	0.161	-1.654	-0.187	0.307
	FFI+Collab	0.217	-0.167	0.601	0.157	1.384	0.157	0.511
Collab	FFI+Collab	0.484	0.100	0.868	0.157	3.084	0.349	0.009

Coher

Within Subjects Effects

Cases	Sum of Squares	df	Mean Square	F	*p*	$η^2$
Time	1.873 [a]	2 [a]	0.937 [a]	3.657 [a]	0.028 [a]	0.016
Time ✻ expgroup	1.663 [a]	4 [a]	0.416 [a]	1.623 [a]	0.171 [a]	0.014
Residuals	38.414	150	0.256			

Between Subjects Effects

Cases	Sum of Squares	df	Mean Square	F	*p*	$η^2$
expgroup	9.808	2	4.904	5.342	0.007	0.081
Residuals	68.846	75	0.918			

Post Hoc Comparisons - Time

			95% CI for Mean Difference					
		Mean Difference	Lower	Upper	SE	t	Cohen's *d*	p_{bonf}
T1	T2	-0.187	-0.384	0.009	0.081	-2.306	-0.261	0.067
	T3	0.006	-0.191	0.202	0.081	0.070	0.008	1.000
T2	T3	0.193	-0.004	0.389	0.081	2.377	0.269	0.056

Post Hoc Comparisons - expgroup

			95% CI for Mean Difference					
		Mean Difference	Lower	Upper	SE	t	Cohen's *d*	p_{bonf}
Comp	Collab	-0.280	-0.663	0.103	0.156	-1.790	-0.203	0.233
	FFI+Collab	0.217	-0.156	0.590	0.152	1.427	0.162	0.474
Collab	FFI+Collab	0.497	0.124	0.870	0.152	3.266	0.370	0.005

Cohes

Within Subjects Effects

Cases	Sum of Squares	df	Mean Square	F	*p*	η²
Time	1.058 [a]	2 [a]	0.529 [a]	2.347 [a]	0.099 [a]	0.011
Time ✻ expgroup	1.134 [a]	4 [a]	0.284 [a]	1.258 [a]	0.289 [a]	0.011
Residuals	33.806	150	0.225			

Between Subjects Effects

Cases	Sum of Squares	df	Mean Square	F	*p*	η²
expgroup	4.480	2	2.240	2.841	0.065	0.045
Residuals	59.131	75	0.788			

Acc

Within Subjects Effects

Cases	Sum of Squares	df	Mean Square	F	*p*	η²
Time	0.146 [a]	2 [a]	0.073 [a]	0.412 [a]	0.663 [a]	0.001
Time ✻ expgroup	2.660 [a]	4 [a]	0.665 [a]	3.763 [a]	0.006 [a]	0.025
Residuals	26.510	150	0.177			

Between Subjects Effects

Cases	Sum of Squares	df	Mean Square	F	*p*	η²
expgroup	3.580	2	1.790	1.877	0.160	0.034
Residuals	71.518	75	0.954			

Post Hoc Comparisons - expgroup ✻ Time

			95% CI for Mean Difference				
		Mean Difference	Lower	Upper	SE	t	*p* bonf
Comp, T1	Collab, T1	-0.240	-0.850	0.370	0.187	-1.286	1.000
	FFI+Collab, T1	0.306	-0.288	0.899	0.182	1.683	1.000
	Comp, T2	0.200	-0.187	0.587	0.119	1.682	1.000
	Collab, T2	-0.120	-0.730	0.490	0.187	-0.643	1.000
	FFI+Collab, T2	-0.051	-0.645	0.542	0.182	-0.283	1.000
	Comp, T3	-0.040	-0.427	0.347	0.119	-0.336	1.000
	Collab, T3	-0.160	-0.770	0.450	0.187	-0.857	1.000
	FFI+Collab, T3	0.091	-0.502	0.685	0.182	0.503	1.000
Collab, T1	FFI+Collab, T1	0.546	-0.048	1.139	0.182	3.005	0.115
	Comp, T2	0.440	-0.170	1.050	0.187	2.357	0.717
	Collab, T2	0.120	-0.267	0.507	0.119	1.009	1.000
	FFI+Collab, T2	0.189	-0.405	0.782	0.182	1.038	1.000
	Comp, T3	0.200	-0.410	0.810	0.187	1.071	1.000
	Collab, T3	0.080	-0.307	0.467	0.119	0.673	1.000
	FFI+Collab, T3	0.331	-0.262	0.925	0.182	1.825	1.000
FFI+Collab, T1	Comp, T2	-0.106	-0.699	0.488	0.182	-0.582	1.000
	Collab, T2	-0.426	-1.019	0.168	0.182	-2.344	0.741
	FFI+Collab, T2	-0.357	-0.723	0.009	0.112	-3.179	0.065
	Comp, T3	-0.346	-0.939	0.248	0.182	-1.903	1.000
	Collab, T3	-0.466	-1.059	0.128	0.182	-2.564	0.413
	FFI+Collab, T3	-0.214	-0.580	0.152	0.112	-1.907	1.000
Comp, T2	Collab, T2	-0.320	-0.930	0.290	0.187	-1.714	1.000
	FFI+Collab, T2	-0.251	-0.845	0.342	0.182	-1.384	1.000
	Comp, T3	-0.240	-0.627	0.147	0.119	-2.018	1.000
	Collab, T3	-0.360	-0.970	0.250	0.187	-1.928	1.000
	FFI+Collab, T3	-0.109	-0.702	0.485	0.182	-0.598	1.000
Collab, T2	FFI+Collab, T2	0.069	-0.525	0.662	0.182	0.378	1.000
	Comp, T3	0.080	-0.530	0.690	0.187	0.429	1.000

Post Hoc Comparisons - expgroup ✻ Time

		Mean Difference	95% CI for Mean Difference Lower	Upper	SE	t	p_{bonf}
	Collab, T3	-0.040	-0.427	0.347	0.119	-0.336	1.000
	FFI+Collab, T3	0.211	-0.382	0.805	0.182	1.164	1.000
FFI+Collab, T2	Comp, T3	0.011	-0.582	0.605	0.182	0.063	1.000
	Collab, T3	-0.109	-0.702	0.485	0.182	-0.598	1.000
	FFI+Collab, T3	0.143	-0.223	0.509	0.112	1.271	1.000
Comp, T3	Collab, T3	-0.120	-0.730	0.490	0.187	-0.643	1.000
	FFI+Collab, T3	0.131	-0.462	0.725	0.182	0.724	1.000
Collab, T3	FFI+Collab, T3	0.251	-0.342	0.845	0.182	1.384	1.000

* p < .05

Mech

Within Subjects Effects

Cases	Sum of Squares	df	Mean Square	F	*p*	η²
Time	1.146 [a]	2 [a]	0.573 [a]	1.969 [a]	0.143 [a]	0.007
Time ✻ expgroup	5.376 [a]	4 [a]	1.344 [a]	4.616 [a]	0.002 [a]	0.034
Residuals	43.675	150	0.291			

Between Subjects Effects

Cases	Sum of Squares	df	Mean Square	F	*p*	η²
expgroup	11.146	2	5.573	4.320	0.017	0.071
Residuals	96.739	75	1.290			

Post Hoc Comparisons - expgroup ✻ Time

			95% CI for Mean Difference					
		Mean Difference	Lower	Upper	SE	t	*p* bonf	
Comp, T1	Collab, T1	-0.320	-1.049	0.409	0.223	-1.432	1.000	
	FFI+Collab, T1	0.639	-0.070	1.347	0.217	2.938	0.139	
	Comp, T2	0.200	-0.297	0.697	0.153	1.310	1.000	
	Collab, T2	-0.160	-0.889	0.569	0.223	-0.716	1.000	
	FFI+Collab, T2	0.353	-0.356	1.062	0.217	1.623	1.000	
	Comp, T3	0.280	-0.217	0.777	0.153	1.835	1.000	
	Collab, T3	0.200	-0.529	0.929	0.223	0.895	1.000	
	FFI+Collab, T3	0.317	-0.392	1.026	0.217	1.459	1.000	
Collab, T1	FFI+Collab, T1	0.959	0.250	1.667	0.217	4.410	< .001	***
	Comp, T2	0.520	-0.209	1.249	0.223	2.327	0.769	
	Collab, T2	0.160	-0.337	0.657	0.153	1.048	1.000	
	FFI+Collab, T2	0.673	-0.036	1.382	0.217	3.095	0.085	
	Comp, T3	0.600	-0.129	1.329	0.223	2.685	0.292	
	Collab, T3	0.520	0.023	1.017	0.153	3.407	0.030	*
	FFI+Collab, T3	0.637	-0.072	1.346	0.217	2.931	0.142	
FFI+Collab, T1	Comp, T2	-0.439	-1.147	0.270	0.217	-2.018	1.000	
	Collab, T2	-0.799	-1.507	-0.090	0.217	-3.674	0.012	*

Post Hoc Comparisons - expgroup ✻ Time

		Mean Difference	95% CI for Mean Difference		SE	t	p_{bonf}
			Lower	Upper			
	FFI+Collab, T2	-0.286	-0.756	0.184	0.144	-1.981	1.000
	Comp, T3	-0.359	-1.067	0.350	0.217	-1.650	1.000
	Collab, T3	-0.439	-1.147	0.270	0.217	-2.018	1.000
	FFI+Collab, T3	-0.321	-0.791	0.148	0.144	-2.229	0.983
Comp, T2	Collab, T2	-0.360	-1.089	0.369	0.223	-1.611	1.000
	FFI+Collab, T2	0.153	-0.556	0.862	0.217	0.703	1.000
	Comp, T3	0.080	-0.417	0.577	0.153	0.524	1.000
	Collab, T3	1.079e -15	-0.729	0.729	0.223	4.829e -15	1.000
	FFI+Collab, T3	0.117	-0.592	0.826	0.217	0.539	1.000
Collab, T2	FFI+Collab, T2	0.513	-0.196	1.222	0.217	2.359	0.708
	Comp, T3	0.440	-0.289	1.169	0.223	1.969	1.000
	Collab, T3	0.360	-0.137	0.857	0.153	2.359	0.706
	FFI+Collab, T3	0.477	-0.232	1.186	0.217	2.195	1.000
FFI+Collab, T2	Comp, T3	-0.073	-0.782	0.636	0.217	-0.335	1.000
	Collab, T3	-0.153	-0.862	0.556	0.217	-0.703	1.000
	FFI+Collab, T3	-0.036	-0.506	0.434	0.144	-0.248	1.000
Comp, T3	Collab, T3	-0.080	-0.809	0.649	0.223	-0.358	1.000
	FFI+Collab, T3	0.037	-0.672	0.746	0.217	0.171	1.000
Collab, T3	FFI+Collab, T3	0.117	-0.592	0.826	0.217	0.539	1.000

* p < .05, ** p < .01, *** p < .001

Lex

Within Subjects Effects

Cases	Sum of Squares	df	Mean Square	F	*p*	η^2
Time	0.224	2	0.112	0.653	0.522	0.002
Time ✻ expgroup	0.619	4	0.155	0.900	0.466	0.006
Residuals	25.783	150	0.172			

Between Subjects Effects

Cases	Sum of Squares	df	Mean Square	F	*p*	η^2
expgroup	7.332	2	3.666	3.949	0.023	0.071
Residuals	69.613	75	0.928			

Post Hoc Comparisons - expgroup

			95% CI for Mean Difference					
		Mean Difference	Lower	Upper	SE	t	Cohen's *d*	p_{bonf}
Comp	Collab	-0.347	-0.732	0.039	0.157	-2.204	-0.249	0.092
	FFI+Collab	0.056	-0.319	0.431	0.153	0.364	0.041	1.000
Collab	FFI+Collab	0.402	0.028	0.777	0.153	2.629	0.298	0.031

* p < .05

APPENDIX D

D1 AQ preliminary item description

	Item	m%	f%	c%	M	95% CI	SD	Skew	Kur	HIc
Writing in Spanish	1	0	1.1	49.5	4.34	[4.17 – 4.51]	0.82	-1.45	2.68	0.43
	2	3.3	5.4	94.6	3.14	[2.91 – 3.37]	1.08	-0.06	-0.76	0.40
	3	1.1	15.1	6.5	2.57	[2.33 – 2.81]	1.14	0.55	-0.54	0.30
	4	2.2	5.4	95.7	3.73	[3.47 – 3.99]	1.22	-0.76	-0.50	0.32
	5	1.1	3.2	96.8	3.76	[3.54 – 3.97]	1.03	-0.62	0.04	0.44
	6	0	5.4	97.8	3.98	[3.79 – 4.17]	0.91	-0.41	-0.79	0.37
	7	1.1	2.2	37.6	4.01	[3.79 – 4.23]	1.04	-0.99	0.35	0.37
	8	0	1.1	24.7	3.89	[3.71 – 4.07]	0.87	-0.60	0.33	0.36
	9	0	0	28	3.86	[3.67 – 4.05]	0.91	-0.25	-0.88	0.44
	10	0	0	22.6	3.75	[3.57 – 3.93]	0.86	0.09	-0.96	0.34
Writing in English	11	0	4.3	10.8	3.09	[2.86 – 3.32]	1.11	0.12	-1.01	0.55
	12	2.2	16.1	7.5	2.66	[2.41 – 2.92]	1.22	0.37	-0.97	0.28
	13	5.5	15.1	8.6	2.94	[2.68 – 3.21]	1.23	-0.12	-0.98	0.46
	14	5.5	2.2	24.7	3.87	[3.66 – 4.08]	0.98	-0.89	0.55	0.44
	15	0	6.5	20.4	3.42	[3.17 – 3.67]	1.20	-0.35	-0.85	0.57
	16	1.1	3.2	9.7	3.43	[3.23 – 3.63]	0.96	-0.51	-0.07	0.43
	17	1.1	2.2	29	3.80	[3.57 – 4.03]	1.08	-0.67	-0.39	0.33
	18	0	18.3	5.4	2.51	[2.27 – 2.74]	1.12	0.50	-0.46	0.47
	19	0	4.3	11.8	3.51	[3.30 – 3.71]	1	-0.69	0.09	0.10
	20	1.1	12.9	9.7	2.99	[2.74 – 3.24]	1.19	-0.10	-0.87	0.58
Collaborative work	21	1.1	2.2	36.6	4.11	[3.92 – 4.30]	0.92	-1.21	1.76	0.61
	22	0	2.2	51.6	4.35	[4.17 – 4.53]	0.87	-1.78	3.87	0.65
	23	1.1	11.8	20.4	3.50	[3.23 – 3.77]	1.26	-0.75	-0.43	0.53
	24	2.2	6.5	15.1	3.44	[3.21 – 3.66]	1.08	-0.51	-0.08	0.37
	25	3.3	24.7	5.4	2.42	[2.17 – 2.67]	1.16	0.47	-0.53	0.48
	26	0	2.2	55.9	4.27	[4.06 – 4.49]	1.03	-1.44	1.36	-0.17
	27	0	7.5	6.5	3.20	[3 – 3.40]	0.97	-0.48	0.33	0.12
	28	0	2.2	29	3.84	[3.62 – 4.05]	1.02	-0.67	-0.09	0.64
Collaborative writing	29	0	9.7	10.8	3.03	[2.80 – 3.27]	1.13	-0.02	-0.61	0.37
	30	1.1	1.1	10.8	3.36	[3.17 – 3.54]	0.86	0.19	-0.05	0.31
	31	1.1	3.2	41.9	4.04	[3.82 – 4.27]	1.09	-1.10	0.52	0.38
	32	0	14	7.5	2.88	[2.63 – 3.12]	1.17	-0.01	-0.88	0.44
	33	2.2	18.3	3.2	2.47	[2.24 – 2.70]	1.09	0.43	-0.61	0.36
	34	0	7.5	9.7	2.93	[2.70 – 3.17]	1.11	0.23	-0.72	0.12
	35	1.1	0	19.4	3.78	[3.60 – 3.95]	0.84	-0.24	-0.52	0.49
	36	0	1.1	20.4	3.60	[3.40 – 3.81]	0.98	-0.15	-0.63	0.47
English learning	37	0	5.4	8.6	3.46	[3.26 – 3.66]	0.97	-0.82	0.52	0.21
	38	1.1	4.3	7.5	3.18	[2.97 – 3.39]	1	-0.16	-0.50	0.48
	39	0	7.5	20.4	3.44	[3.19 – 3.69]	1.21	-0.45	-0.75	0.04
	40	0	3.2	38.7	4.09	[3.87 – 4.30]	1.03	-1.37	1.53	0.23
	41	1.1	9.7	12.9	2.97	[2.71 – 3.22]	1.22	0.18	-1.01	-0.04
	42	1.1	14	17.2	2.91	[2.63 – 3.19]	1.34	0.25	-1.16	0.16
	43	1.1	0	73.1	4.70	[4.58 – 4.82]	0.57	-1.77	2.18	0.24
	44	0	12.9	16.1	3.13	[2.86 – 3.40]	1.29	-0.16	-1.08	0.12

Legend: Item = question number; m% = percentage of missing values; f% = percentage of cases with a floor value; c% = percentage of cases with a ceiling value; M = mean; 95% CI = Upper and lower limit of Confidence Interval; SD = standard deviation; Skew = skewness; Kur = Kurtosis; HIc: Corrected homogeneity index (within each attitudinal area)

D2 AQ item intercorrelation indices (Spearman)

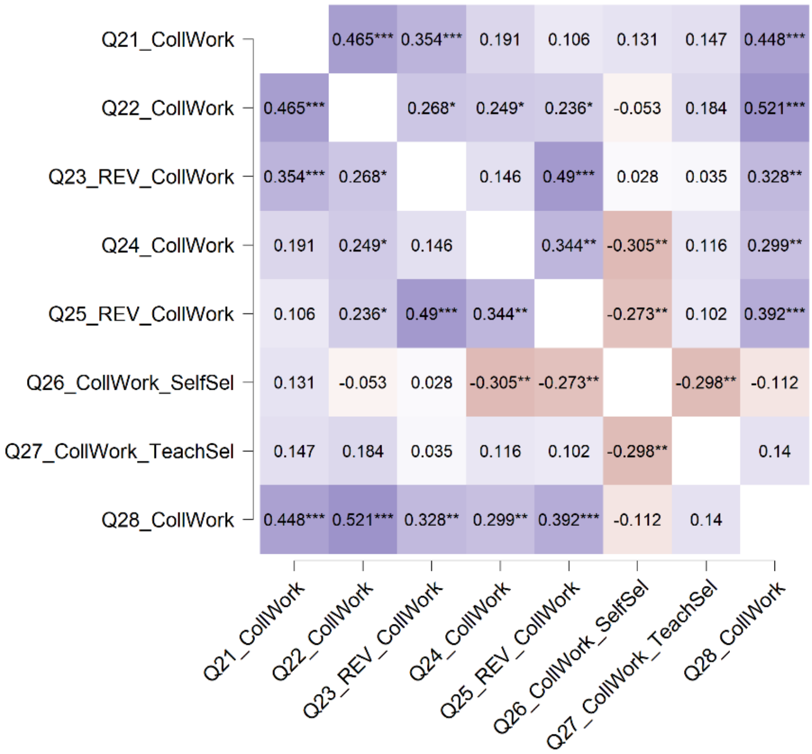
L1 writing

	Q1_L1_Anx	Q2_REV_L1_Anx	Q3_REV_L1_Anx	Q4_REV_L1_Anx	Q5_REV_L1_Anx	Q6_L1_SelfEff	Q7_L1_SelfEff	Q8_L1_SelfEff	Q9_L1_SelfEff	Q10_L1_SelfEff
Q1_L1_Anx		0.306**	0.074	0.179	0.188	0.432***	0.182	0.215*	0.191	0.274**
Q2_REV_L1_Anx	0.306**		0.321**	0.293**	0.235*	0.248*	-0.068	0.063	0.333**	-0.001
Q3_REV_L1_Anx	0.074	0.321**		0.085	0.122	0.061	0.137	0.049	0.12	0.159
Q4_REV_L1_Anx	0.179	0.293**	0.085		0.456***	0.096	-0.013	0.155	0.215*	0.134
Q5_REV_L1_Anx	0.188	0.235*	0.122	0.456***		0.148	0.034	0.362***	0.227*	0.241*
Q6_L1_SelfEff	0.432***	0.248*	0.061	0.096	0.148		0.239*	0.055	0.17	0.265*
Q7_L1_SelfEff	0.182	-0.068	0.137	-0.013	0.034	0.239*		0.243*	0.275**	0.186
Q8_L1_SelfEff	0.215*	0.063	0.049	0.155	0.362***	0.055	0.243*		0.28**	0.262*
Q9_L1_SelfEff	0.191	0.333**	0.12	0.215*	0.227*	0.17	0.275**	0.28**		0.126
Q10_L1_SelfEff	0.274**	-0.001	0.159	0.134	0.241*	0.265*	0.186	0.262*	0.126	

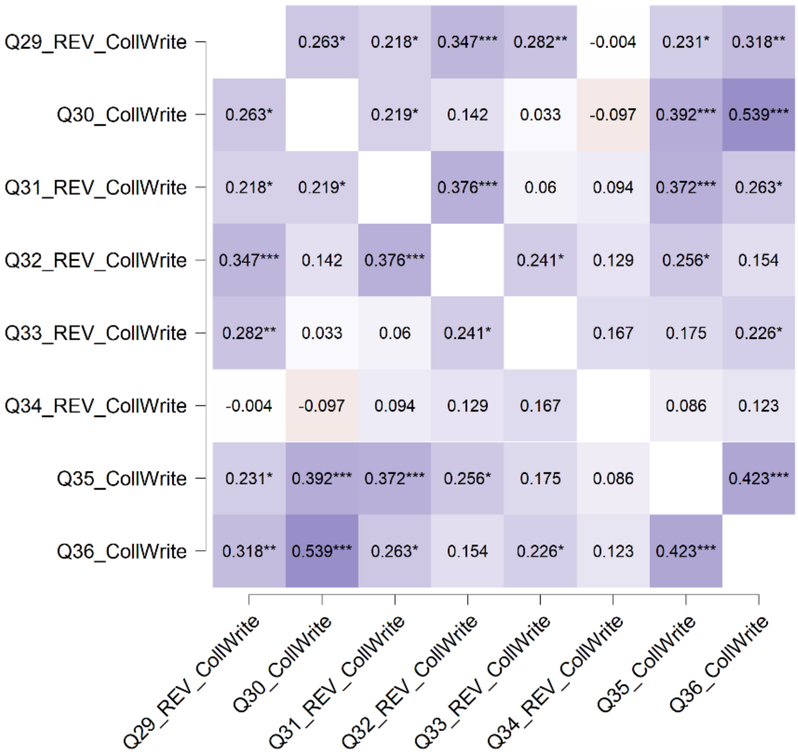
L2 Writing

	Q11_REV_L2_Anx	Q12_L1_Anx	Q13_L2_Anx	Q14_L2_Anx	Q15_REV_L2_Anx	Q16_L2_SelfEff	Q17_L2_SelfEff	Q18_REV_L2_SelfEff	Q19_REV_L2_SelfEff
Q11_REV_L2_Anx		0.206	0.221*	0.176	0.514***	0.369***	0.168	0.373***	0.163
Q12_L1_Anx	0.206		0.385***	0.123	0.134	0.085	0.255*	0.132	-0.008
Q13_L2_Anx	0.221*	0.385***		0.369***	0.259*	0.159	0.304**	0.318**	-0.154
Q14_L2_Anx	0.176	0.123	0.369***		0.424***	0.265*	0.236*	0.284**	-0.013
Q15_REV_L2_Anx	0.514***	0.134	0.259*	0.424***		0.335**	0.306**	0.358***	0.142
Q16_L2_SelfEff	0.369***	0.085	0.159	0.265*	0.335**		0.186	0.282**	0.146
Q17_L2_SelfEff	0.168	0.255*	0.304**	0.236*	0.306**	0.186		0.05	-0.047
Q18_REV_L2_SelfEff	0.373***	0.132	0.318**	0.284**	0.358***	0.282**	0.05		0.109
Q19_REV_L2_SelfEff	0.163	-0.008	-0.154	-0.013	0.142	0.146	-0.047	0.109	

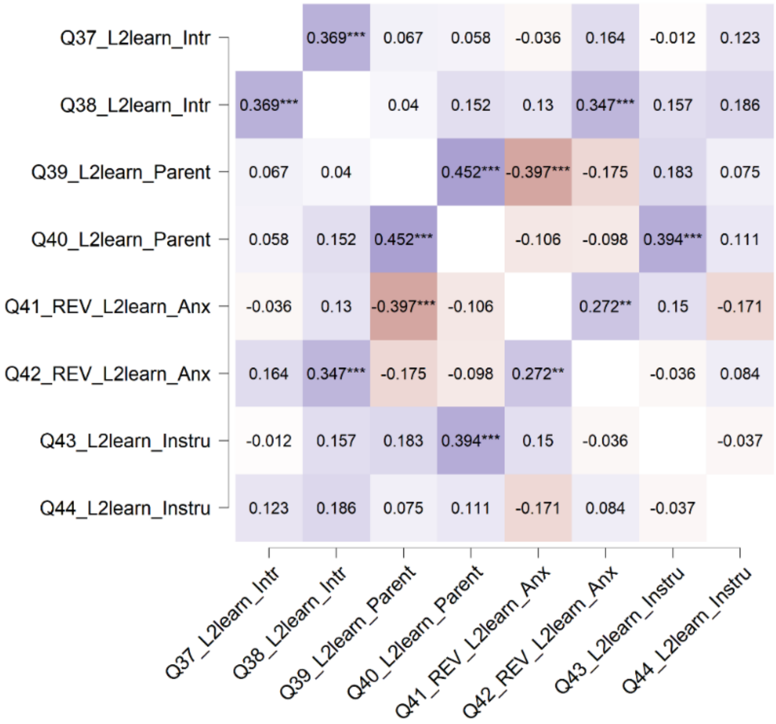
Collaborative work



Collaborative writing



English learning



D3 AQ multidimensional Principal Component Analysis

1st Round

Total Variance Explained

	Initial Eigenvalues			Extraction Sums of Squared Loadings		
Component	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	4.555	15.708	15.708	4.555	15.708	15.708
2	4.224	14.565	30.273	4.224	14.565	30.273
3	2.030	7.001	37.274	2.030	7.001	37.274
4	1.885	6.499	43.772	1.885	6.499	43.772
5	1.604	5.531	49.303	1.604	5.531	49.303
6	1.498	5.164	54.467	1.498	5.164	54.467
7	1.321	4.557	59.024	1.321	4.557	59.024
8	1.204	4.151	63.175	1.204	4.151	63.175
9	1.142	3.937	67.112	1.142	3.937	67.112
10	1.019	3.513	70.625	1.019	3.513	70.625
11	.933	3.216	73.842			
12	.891	3.071	76.912			
13	.742	2.558	79.471			
14	.669	2.306	81.777			
15	.608	2.097	83.874			
16	.586	2.022	85.896			
17	.567	1.956	87.852			
18	.539	1.859	89.711			
19	.468	1.613	91.324			
20	.434	1.495	92.819			
21	.353	1.219	94.038			
22	.320	1.104	95.141			
23	.295	1.017	96.159			
24	.249	.860	97.019			
25	.228	.787	97.806			
26	.200	.689	98.494			
27	.181	.624	99.118			
28	.138	.475	99.594			
29	.118	.406	100.000			

Extraction Method: Principal Component Analysis.

Unrotated Component Matrix[a]

	Component									
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
Q14_L2_Anx	.592	-.274	-.180	-.094	.171	-.303	.123	-.209	-.198	-.291
Q9_L1_SelfEff	.535	-.030	.106	-.333	-.186	.030	-.190	-.167	-.287	.018
Q15_REV_L2_Anx	.523	-.356	-.402	.164	.024	-.020	-.060	-.143	-.165	.186
Q4_REV_L1_Anx	.518	-.086	-.462	-.184	-.023	-.262	.110	-.196	.011	-.215
Q13_L2_Anx	.504	-.379	.176	.012	.068	-.314	.375	.038	.143	-.060
Q17_L2_SelfEff	.497	-.200	.415	-.111	-.166	.282	.124	-.439	-.106	.174
Q5_REV_L1_Anx	.494	-.060	-.377	-.177	-.289	-.195	-.224	.188	.128	-.089
Q18_REV_L2_SelfEff	.457	-.268	-.284	.424	.058	.067	-.113	.101	.246	-.229
Q8_L1_SelfEff	.438	.001	.224	.177	-.365	-.032	-.400	.262	.109	-.038
Q23_REV_CollWork	.261	.687	-.035	.244	-.174	.103	-.212	.075	-.130	.066
Q25_REV_CollWork	.411	.616	-.035	-.015	-.150	.246	.276	.042	.113	-.066
Q30_CollWrite	.083	.580	.035	.046	.429	-.210	-.086	-.056	.067	.251
Q35_CollWrite	.382	.534	-.061	-.094	.302	.026	-.077	-.217	-.203	-.061
Q28_CollWork	.410	.510	.338	.166	-.093	-.227	.251	.133	-.027	-.129
Q29_REV_CollWrite	.138	.506	-.204	.431	-.227	-.037	.021	-.293	.042	.300
Q36_CollWrite	.335	.492	-.090	-.079	.436	-.216	.079	-.070	.195	.261
Q20_REV_L2_SelfEff	.367	-.479	.012	.410	.189	.283	.051	.047	-.106	-.024
Q3_REV_L1_Anx	.216	-.475	.083	-.333	-.307	.063	.357	.163	.182	.268
Q32_REV_CollWrite	.417	.439	-.136	-.038	-.437	-.025	-.095	-.122	.034	-.029
Q11_REV_L2_Anx	.349	-.433	-.239	.422	-.042	.237	.062	.241	-.281	.169
Q7_L1_SelfEff	.466	-.023	.559	-.225	-.042	.317	-.193	-.214	.058	-.111
Q2_REV_L1_Anx	.384	-.028	-.162	-.535	.034	-.049	.170	.393	-.194	.352
Q12_L1_Anx	.215	-.178	.417	.469	-.060	-.192	.328	-.096	.113	.178
Q16_L2_SelfEff	.371	-.312	-.046	.134	.435	.328	-.057	-.071	.109	.002
Q33_REV_CollWrite	.175	.387	-.384	-.070	-.044	.459	.201	.052	.443	.098
Q24_CollWork	.369	.342	.159	-.197	.312	.378	.126	.276	.068	-.349
Q6_L1_SelfEff	.356	-.235	.116	-.165	.284	.049	-.440	.152	.014	.323
Q21_CollWork	.351	.372	.338	.233	.077	-.226	.028	.424	-.285	-.027
Q10_L1_SelfEff	.313	-.310	.259	-.015	.061	-.322	-.317	-.028	.534	.034

Extraction Method: Principal Component Analysis.
a. 10 components extracted.

2nd Round

Pattern Matrix[a]

	Component	
	1	**2**
Q14_L2_Anx	.648	.053
Q15_REV_L2_Anx	.633	-.052
Q13_L2_Anx	.628	-.082
Q20_REV_L2_SelfEff	.565	-.236
Q18_REV_L2_SelfEff	.530	-.008
Q17_L2_SelfEff	.529	.071
Q11_REV_L2_Anx	.525	-.205
Q4_REV_L1_Anx	.486	.181
Q16_L2_SelfEff	.480	-.089
Q9_L1_SelfEff	.471	.238
Q5_REV_L1_Anx	.452	.191
Q3_REV_L1_Anx	.434	-.307
Q10_L1_SelfEff	.430	-.116
Q6_L1_SelfEff	.427	-.029
Q7_L1_SelfEff	.409	.210
Q8_L1_SelfEff	.372	.217
Q2_REV_L1_Anx	.341	.165
Q12_L1_Anx	.276	-.049
Q25_REV_CollWork	.025	.739
Q23_REV_CollWork	-.140	.727
Q35_CollWrite	.044	.654
Q28_CollWork	.080	.646
Q36_CollWrite	.025	.593
Q32_REV_CollWrite	.123	.588
Q30_CollWrite	-.235	.546
Q29_REV_CollWrite	-.150	.509
Q21_CollWork	.103	.498
Q24_CollWork	.134	.480
Q33_REV_CollWrite	-.055	.424

Extraction Method: Principal Component Analysis.
Rotation Method: Oblimin with Kaiser Normalization.[a]
a. Rotation converged in 6 iterations.

3rd Round

Pattern Matrix[a]

	Component	
	1	**2**
Q14_L2_Anx	.656	.042
Q15_REV_L2_Anx	.638	-.063
Q13_L2_Anx	.611	-.087
Q20_REV_L2_SelfEff	.552	-.241
Q18_REV_L2_SelfEff	.531	-.016
Q17_L2_SelfEff	.517	.066
Q11_REV_L2_Anx	.517	-.211
Q4_REV_L1_Anx	.504	.170
Q9_L1_SelfEff	.485	.228
Q16_L2_SelfEff	.481	-.096
Q5_REV_L1_Anx	.470	.181
Q6_L1_SelfEff	.436	-.037
Q3_REV_L1_Anx	.433	-.314
Q10_L1_SelfEff	.419	-.119
Q7_L1_SelfEff	.407	.205
Q8_L1_SelfEff	.366	.214
Q2_REV_L1_Anx	.364	.154
Q25_REV_CollWork	.029	.739
Q23_REV_CollWork	-.133	.729
Q35_CollWrite	.057	.651
Q28_CollWork	.067	.649
Q36_CollWrite	.034	.592
Q32_REV_CollWrite	.138	.584
Q30_CollWrite	-.234	.551
Q29_REV_CollWrite	-.154	.513
Q21_CollWork	.091	.500
Q24_CollWork	.148	.476
Q33_REV_CollWrite	-.038	.421

Extraction Method: Principal Component Analysis.
Rotation Method: Oblimin with Kaiser Normalization.
a. Rotation converged in 7 iterations.

Structure Matrix

	Component	
	1	2
Q14_L2_Anx	.658	.070
Q15_REV_L2_Anx	.635	-.035
Q13_L2_Anx	.607	-.060
Q20_REV_L2_SelfEff	.542	-.218
Q18_REV_L2_SelfEff	.531	.007
Q17_L2_SelfEff	.520	.089
Q4_REV_L1_Anx	.511	.191
Q11_REV_L2_Anx	.508	-.189
Q9_L1_SelfEff	.495	.249
Q5_REV_L1_Anx	.477	.201
Q16_L2_SelfEff	.477	-.076
Q6_L1_SelfEff	.434	-.018
Q3_REV_L1_Anx	.420	-.295
Q7_L1_SelfEff	.415	.222
Q10_L1_SelfEff	.414	-.101
Q8_L1_SelfEff	.375	.229
Q2_REV_L1_Anx	.371	.170
Q25_REV_CollWork	.061	.740
Q23_REV_CollWork	-.102	.723
Q35_CollWrite	.085	.654
Q28_CollWork	.095	.652
Q36_CollWrite	.059	.594
Q32_REV_CollWrite	.163	.590
Q30_CollWrite	-.211	.541
Q29_REV_CollWrite	-.132	.507
Q21_CollWork	.112	.504
Q24_CollWork	.169	.482
Q33_REV_CollWrite	-.020	.419

Extraction Method: Principal Component Analysis.
Rotation Method: Oblimin with Kaiser Normalization.

Component Correlation Matrix

Component	1	2
1	1.000	.043
2	.043	1.000

Extraction Method: Principal Component Analysis.
Rotation Method: Oblimin with Kaiser Normalization.

Final internal scale validation

Attitudes towards writing

Item-Total Statistics

	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation (HIc)	Cronbach's Alpha if Item Deleted
Q2_REV_L1_Anx	55.56	83.027	.314	.837
Q3_REV_L1_Anx	56.07	81.870	.365	.834
Q4_REV_L1_Anx	55.04	78.957	.447	.830
Q5_REV_L1_Anx	54.90	81.310	.467	.829
Q6_L1_SelfEff	54.70	83.325	.391	.833
Q7_L1_SelfEff	54.74	82.445	.356	.835
Q8_L1_SelfEff	54.71	84.763	.354	.834
Q9_L1_SelfEff	54.82	82.982	.415	.832
Q10_L1_SelfEff	54.90	84.060	.346	.835
Q11_REV_L2_Anx	55.58	79.637	.490	.827
Q13_L2_Anx	55.71	77.486	.521	.825
Q14_L2_Anx	54.82	80.065	.532	.826
Q15_REV_L2_Anx	55.29	75.597	.623	.819
Q16_L2_SelfEff	55.33	81.863	.441	.830
Q17_L2_SelfEff	54.90	79.838	.496	.827
Q18_REV_L2_SelfEff	56.15	79.908	.453	.829
Q20_REV_L2_SelfEff	55.73	78.979	.486	.828

Attitudes towards collaboration

Item-Total Statistics

	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation (HIc)	Cronbach's Alpha if Item Deleted
Q21_CollWork	32.15	39.853	.379	.801
Q23_REV_CollWork	32.80	34.610	.602	.778
Q24_CollWork	32.78	38.850	.405	.799
Q25_REV_CollWork	33.83	34.995	.643	.774
Q28_CollWork	32.37	37.111	.575	.783
Q29_REV_CollWrite	33.27	38.725	.379	.803
Q30_CollWrite	32.95	40.423	.385	.801
Q32_REV_CollWrite	33.36	37.683	.448	.796
Q33_REV_CollWrite	33.77	38.457	.399	.801
Q35_CollWrite	32.49	39.128	.493	.792
Q36_CollWrite	32.70	38.611	.470	.793

AQ Total internal reliability

Reliability Statistics

Cronbach's Alpha	N of Items
.792	28

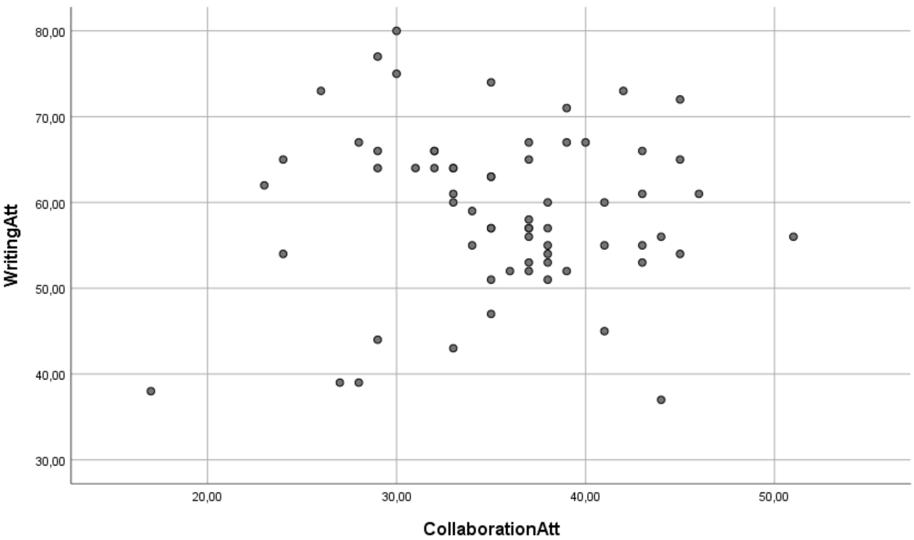
Item-Total Statistics

	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation (HIc)	Cronbach's Alpha if Item Deleted
Q2_REV_L1_Anx	91.47	120.438	.300	.787
Q3_REV_L1_Anx	91.97	124.522	.122	.795
Q4_REV_L1_Anx	90.94	116.704	.389	.782
Q5_REV_L1_Anx	90.83	120.849	.338	.785
Q6_L1_SelfEff	90.67	120.779	.385	.784
Q7_L1_SelfEff	90.70	115.384	.520	.776
Q8_L1_SelfEff	90.67	122.133	.349	.785
Q9_L1_SelfEff	90.77	118.640	.492	.779
Q10_L1_SelfEff	90.83	122.264	.295	.787
Q11_REV_L2_Anx	91.53	122.868	.194	.792
Q13_L2_Anx	91.61	114.950	.463	.778
Q14_L2_Anx	90.74	119.056	.411	.782
Q15_REV_L2_Anx	91.21	115.893	.442	.779
Q16_L2_SelfEff	91.26	120.779	.330	.785
Q17_L2_SelfEff	90.83	114.018	.578	.773
Q18_REV_L2_SelfEff	92.11	119.850	.306	.786
Q20_REV_L2_SelfEff	91.68	121.020	.253	.789
Q21_CollWork	90.55	122.098	.293	.787
Q23_REV_CollWork	91.29	121.593	.199	.793
Q24_CollWork	91.23	124.332	.157	.793
Q25_REV_CollWork	92.30	120.368	.289	.787
Q28_CollWork	90.76	120.925	.318	.786
Q29_REV_CollWrite	91.76	124.833	.114	.796
Q30_CollWrite	91.41	127.045	.061	.795
Q32_REV_CollWrite	91.85	121.731	.259	.789
Q33_REV_CollWrite	92.23	124.732	.114	.796
Q35_CollWrite	90.89	121.481	.373	.784
Q36_CollWrite	91.11	119.604	.376	.783

D4 AQ subscale score calculation

Correlations

		WritingAtt	CollaborationAtt
WritingAtt	Pearson Correlation	1	-.004
	Sig. (2-tailed)		.972
	N	73	66
CollaborationAtt	Pearson Correlation	-.004	1
	Sig. (2-tailed)	.972	
	N	66	81



D5 AQ mean scores one-way ANOVA

WritingAtt

ANOVA - WritingAttMEAN

Cases	Sum of Squares	df	Mean Square	F	*p*	η²
expgroup	3.570	2	1.785	6.587	0.002	0.158
Residuals	18.969	70	0.271			

Post Hoc Comparisons - expgroup

			95% CI for Mean Difference				
		Mean Difference	**Lower**	**Upper**	**SE**	**t**	***p* bonf**
Comp	Collab	7.494e -16	-0.368	0.368	0.154	4.882e -15	1.000
	FFI+Collab	0.458	0.104	0.812	0.148	3.101	0.008
Collab	FFI+Collab	0.458	0.104	0.812	0.148	3.101	0.008

Collab Att

ANOVA - CollabAttMEAN

Cases	Sum of Squares	df	Mean Square	F	*p*	η²
expgroup	0.993	2	0.496	1.341	0.267	0.033
Residuals	28.867	78	0.370			

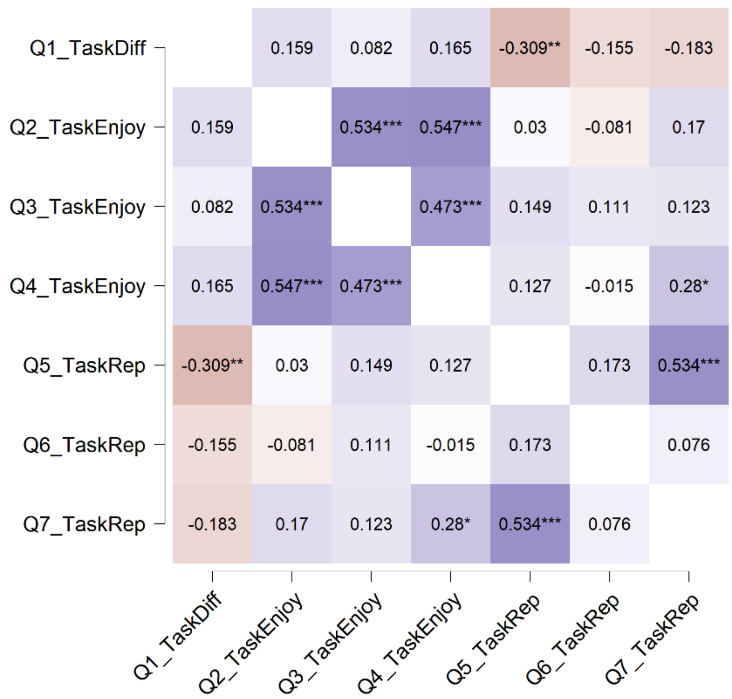
D6 DQ preliminary item description

	Item	m%	f%	c%	M	95% CI	SD	Skew	Kur	HIc
Common	1	0	8.3	11.9	3.14	[2.85 – 3.37]	1.17	-0.15	-0.95	-0.08
	2	1.2	10.7	10.7	3.10	[2.82 – 3.33]	1.15	-0.19	-0.68	0.52
	3	1.2	10.7	10.7	3.17	[2.91 – 3.44]	1.19	-0.33	-0.84	0.56
	4	1.2	15.5	9.5	3.05	[2.79 – 3.34]	1.25	-0.29	-1.02	0.55
	5	0	6	19	3.49	[3.22 – 3.71]	1.10	-0.47	-0.29	0.21
	6	0	1.2	41.7	4.15	[3.94 – 4.33]	0.87	-0.87	0.68	0.04
	7	0	9.5	27.4	3.60	[3.32 – 3.86]	1.22	-0.67	-0.29	0.33
Ind.	8	0	15.4	7.7	2.92	[2.37 – 3.29]	1.13	-0.20	-0.24	0.23
	9	0	19.2	8.3	2.58	[2.09 – 3.08]	1.21	0.47	-0.57	0.30
	10	7.7	8.3	8.3	3.17	[2.67 – 3.66]	1.17	-0.35	-0.92	0.23
	11	0	7.7	26.9	3.73	[3.16 – 4.09]	1.17	-0.92	0.93	-0.06
Collab.	12	0	1.7	25.9	3.56	[3.27 – 3.86]	1.08	-0.12	-0.89	0.30
	13	0	6.9	39.7	4.04	[3.78 – 4.29]	0.94	-0.63	0.56	0.70
	14	3.4	7.1	50	4.33	[4.09 – 4.56]	0.86	-1.42	1.68	0.69
	15	1.7	3.5	42.1	4.05	[3.77 – 4.34]	1.06	-1.17	0.99	0.51
FFI	16	3.2	3.2	12.9	3.43	[3.08 – 3.78]	0.93	-0.20	0.49	--

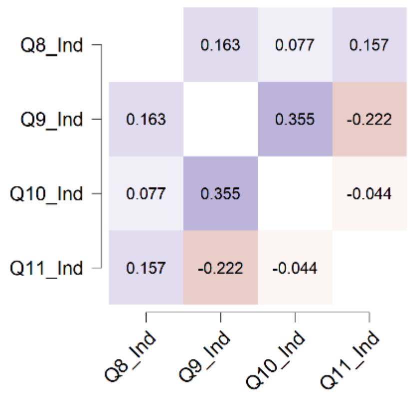
Legend: Ind. = Individual setting; Collab. = Collaborative setting; FFI = Form-focused instruction; Item = question number; m% = percentage of missing values; f% = percentage of cases with a floor value; c% = percentage of cases with a ceiling value; M = mean; 95% CI = Upper and lower limit of Confidence Interval; SD = standard deviation; Skew = skewness; Kur = Kurtosis; HIc: Corrected homogeneity index (within each attitudinal area)

D7 DQ item correlation indices (Spearman)

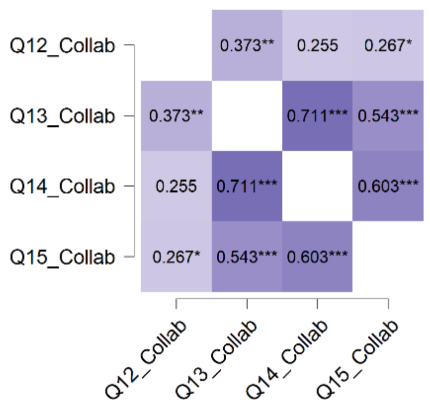
Common block



Individual setting block



Collaborative setting block



D8 DQ unidimensional Principal Component Analysis

Total Variance Explained

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings			Rotation Sums of Squared Loadings[a]
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %	Total
1	2.330	46.604	46.604	2.330	46.604	46.604	2.201
2	1.264	25.274	71.878	1.264	25.274	71.878	1.593
3	.605	12.103	83.981				
4	.420	8.403	92.384				
5	.381	7.616	100.000				

Extraction Method: Principal Component Analysis.

a. When components are correlated. sums of squared loadings cannot be added to obtain a total variance.

Pattern Matrix[a]

	Component	
	1	2
Q2_TaskEnjoy	.880	-.063
Q3_TaskEnjoy	.825	-.009
Q4_TaskEnjoy	.804	.085
Q5_TaskRep	-.074	.882
Q7_TaskRep	.090	.827

Extraction Method: Principal Component Analysis.
Rotation Method: Oblimin with Kaiser Normalization.[a]
a. Rotation converged in 3 iterations.

Structure Matrix

	Component	
	1	2
Q2_TaskEnjoy	.865	.142
Q4_TaskEnjoy	.823	.272
Q3_TaskEnjoy	.823	.183
Q5_TaskRep	.132	.865
Q7_TaskRep	.282	.847

Extraction Method: Principal Component Analysis.
Rotation Method: Oblimin with Kaiser Normalization.

Component Correlation Matrix

Component	1	2
1	1.000	.233
2	.233	1.000

Extraction Method: Principal Component Analysis.
Rotation Method: Oblimin with Kaiser Normalization.

Final internal scale validation

DG task preferences

Reliability Statistics

Cronbach's Alpha	N of Items
.789	3

Item-Total Statistics

	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Cronbach's Alpha if Item Deleted
Q2_TaskEnjoy	6.23	4.507	.663	.678
Q3_TaskEnjoy	6.14	4.519	.617	.726
Q4_TaskEnjoy	6.25	4.413	.609	.736

Perception of TR

Reliability Statistics

Cronbach's Alpha	N of Items
.637	2

Item-Total Statistics

	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Cronbach's Alpha if Item Deleted
Q5_TaskRep	3.60	1.497	.470	.
Q7_TaskRep	3.49	1.217	.470	.

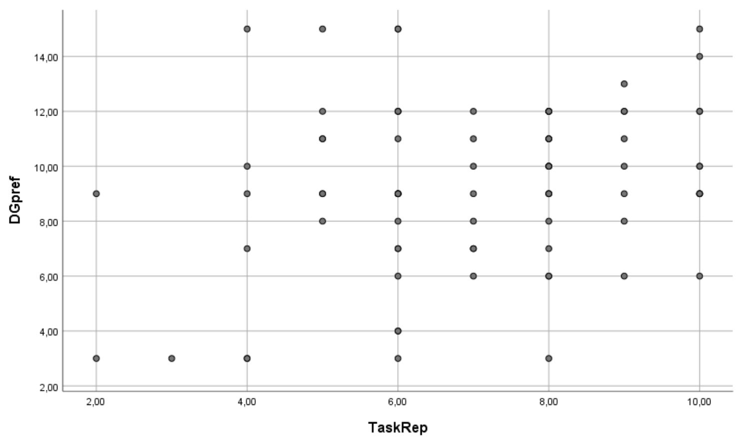
D9 DQ subscale calculation

DG Preferences vs TR

Correlations

		DGpref	TaskRep
DGpref	Pearson Correlation	1	.263*
	Sig. (2-tailed)		.018
	N	81	81
TaskRep	Pearson Correlation	.263*	1
	Sig. (2-tailed)	.018	
	N	81	84

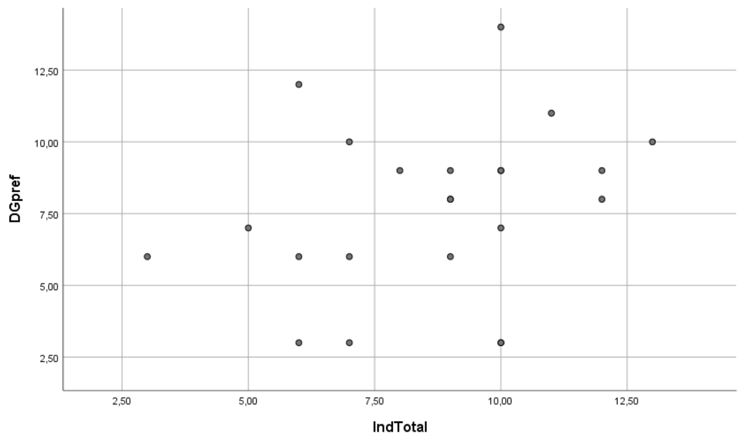
*. Correlation is significant at the 0.05 level (2-tailed).



DG Preferences vs Ind Setting

Correlations

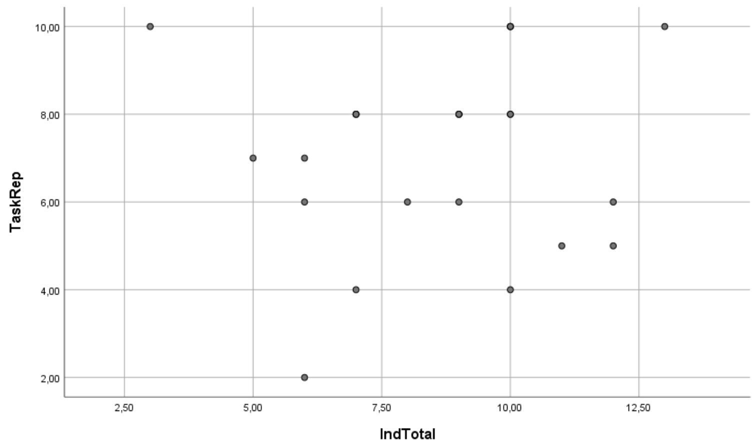
		DGpref	IndTotal
DGpref	Pearson Correlation	1	.285
	Sig. (2-tailed)		.188
	N	25	23
IndTotal	Pearson Correlation	.285	1
	Sig. (2-tailed)	.188	
	N	23	24



TR vs Ind Setting

Correlations

		TaskRep	IndTotal
TaskRep	Pearson Correlation	1	.080
	Sig. (2-tailed)		.710
	N	26	24
IndTotal	Pearson Correlation	.080	1
	Sig. (2-tailed)	.710	
	N	24	24

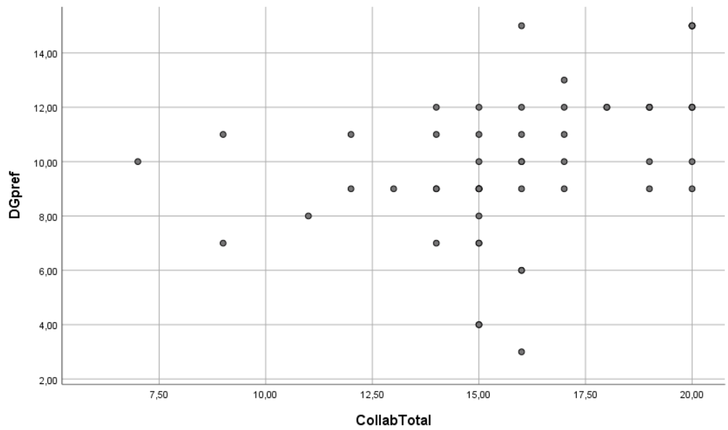


DG Preferences vs Collab Setting

Correlations

		DGpref	CollabTotal
DGpref	Pearson Correlation	1	.413**
	Sig. (2-tailed)		.002
	N	56	54
CollabTotal	Pearson Correlation	.413**	1
	Sig. (2-tailed)	.002	
	N	54	55

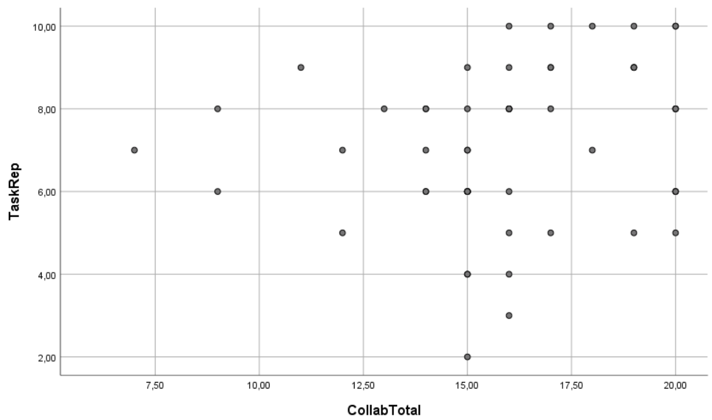
**. Correlation is significant at the 0.01 level (2-tailed).



Task preferences vs Collab Setting

Correlations

		TaskRep	**CollabTotal**
TaskRep	Pearson Correlation	1	.185
	Sig. (2-tailed)		.175
	N	58	55
CollabTotal	Pearson Correlation	.185	1
	Sig. (2-tailed)	.175	
	N	55	55



D10 DQ mean scores one-way ANOVA

DG task preferences

ANOVA - DGprefMEAN

Cases	Sum of Squares	df	Mean Square	F	*p*	η²
expgroup	21.654	2	10.827	14.223	< .001	0.267
Residuals	59.377	78	0.761			

Post Hoc Comparisons - expgroup

		Mean Difference	95% CI for Mean Difference Lower	95% CI for Mean Difference Upper	SE	t	Cohen's *d*	*p* bonf	
Comp	Collab	-1.288	-1.872	-0.704	0.244	-5.271	-1.577	< .001	***
	FFI+Collab	-0.493	-1.058	0.071	0.236	-2.088	-0.507	0.120	
Collab	FFI+Collab	0.795	0.236	1.353	0.234	3.400	0.976	0.003	**

* p < .05, ** p < .01, *** p < .001

Perception of TR

ANOVA - TaskRep

Cases	Sum of Squares	df	Mean Square	F	*p*	η²
expgroup	1.819	2	0.910	0.224	0.800	0.006
Residuals	328.597	81	4.057			

Perceived task difficulty

ANOVA - Q1_TaskDiff

Cases	**Sum of Squares**	**df**	**Mean Square**	**F**	***p***	**η²**
expgroup	1.453	2	0.727	0.522	0.596	0.013
Residuals	112.833	81	1.393			

Collaborative setting

ANOVA - CollabMEAN

Cases	**Sum of Squares**	**df**	**Mean Square**	**F**	***p***	**η²**
expgroup	7.135	1	7.135	16.584	< .001	0.238
Residuals	22.802	53	0.430			

Post Hoc Comparisons - expgroup

		95% CI for Mean Difference						
	Mean Difference	**Lower**	**Upper**	**SE**	**t**	**Cohen's *d***	***p* bonf**	
Collab	FFI+Collab 0.723	0.367	1.080	0.178	4.072	1.103	< .001	***

*** p < .001

D11 Example of focus-group interview transcript

@Begin
@Language: eng, eus, spa
@School: mariturri
@Researcher: Asier
@Participants: *CHI1 MAR001, *CHI2 MAR014, *CHI3 MAR016, *CHI4 MAR012, *CHI5 MAR018, *CHI6 MAR009

(9:03 mins)

*RES: bueno, ¿qué tal entonces? ¿todo bien?
*CHI4: ¡sí!
*RES: vale, entonces yo os voy a ir preguntando y luego para… [the last participant enters the class] venga hombre! venga!
*CHI4: ¡corre!
*CHI5: no grites.
*RES: no gritéis, que está grabando la cámara. Os voy a ir preguntando cosas sobre el experimento, sobre lo que hemos estado haciendo estas semanas, ¿vale? entonces vais a ir contando vuestra opinión decidme la verdad, no me engañéis…
*CHI4: ¿sabes lo que voy a decir? ¡ha sido un chasco!
*RES: podéis decir eso si creéis que ha sido así, y para contestar vais levantando la mano y creara un poquito de debate. Entonces lo primero, ¿cuál creéis que fue el objetivo de las tareas las tareas? Os acordáis que eran "A celebration" y la otra que la tengo también por aquí… ¿primero CHI4 creo que ha sido no? venga.
*CHI2: segundo yo.
*CHI4: para ver qué nivel tenemos de… para ver qué nivel de inglés tenemos.
*RES: sí, vale muy bien.
*CHI2: para aprender a trabajar en parejas.
*CHI2: para ver como trabajamos en parejas en inglés.
*RES: perfecto muy bien (,) ¿CHI3?
*CHI3: ¿para ver cuánta información recogemos?
*RES: vale también puede ser (,) ¿o sea cosas de memoria? ¿cuánto recordáis?
*CHI3: sí.
*RES: vale, ¿algo más?
*CHI1: cuanto sabemos sobre inglés.
*RES: cuanto sabéis sobre inglés…
*CHI5: para practicar.
*RES: para practicar, ¿y exactamente qué? ¿qué podría ser?
*CHI5: lo de…
*CHI6: *ulermena* [comprensión].
*RES: *ulermena.*
*CHI4: gramática.
*CHI1: las frases… o sea pasado presente futuro.
*CHI6: signos de puntuación.
*RES: ¿vale un poco todo, no? vale ¿y qué os han parecido esas dos tareas, las que hicisteis en parejas?
*CHI4: bueno, divertido.

*RES:	¿sí?
*CHI4:	si con la pareja te lo pasas bien y haces chistes.
*CHI2:	¡difícil!
*RES:	difícil dice por aquí (,) ¿te pareció difícil?
*CHI2:	más o menos.
*RES:	vosotros dos trabajasteis juntos? [CHI1 and CHI2 nod] vale.
*CHI4:	a mí la segunda me pareció más difícil que la primera porque yo era la que más cosas apuntaba y en la primera parecía que lo tenía todo y la segunda fue más difícil… bastante más difícil.
*RES:	¿o sea que el segundo día más difícil que el primero? ¿todos así?
*CHI6:	yo el segundo día no (,) el segundo día fue más fácil.
*CHI5:	eso es verdad.
*CHI6:	el segundo día entendí más.
*CHI1:	yo también.
*RES:	aja, tú también, ¿y tú CHI2?
*CHI2:	también.
*CHI4:	el primero me lo sabía de memoria.
*RES:	CHI3, y tú que trabajaste con XXX, que fue más…?
*CHI3:	siempre tenía que elegir su opinión.
*RES:	ah eso, eso os lo voy a preguntar también (,) ¿entonces cómo os sentisteis al hacerlas? ¿os gustó trabajar con vuestra pareja o no? ¿hubieseis cambiado? aquí hay algunos que trabajaron juntos así que estáis un poco coaccionados.
*CHI4:	yo me lo pase bien trabajando con mi pareja porque muchas veces me lo paso bien pero hay en otras ocasiones que me gustaría también estar con otras personas.
*RES:	¿o sea que hubieseis preferido cambiar igual el segundo día con otra persona? a ver CHI2.
*CHI2:	el segundo día cambiar, sí.
*RES:	para que no fuera todo el rato tan tan lo mismo.
*CHI4:	yo estuve bien pero igual si volvieras cambiar de pareja.
*CHI1:	yo lo mismo.
*RES:	tú lo mismo (,) vale (,) CHI3?
*CHI3:	también.
*RES:	también (,) ¿pero estuviste a gusto?
[CHI3 shakes her head]
*RES:	no estuviste a gusto? ¿por qué? a ver dime por qué (,) esto no va a salir de aquí tranquila (,) porque has dicho lo de la opinión… ¿o tú no podías dar tu opinión?
*CHI3:	siempre decía algo pero apuntaba ella todo.
*RES:	uhum vale (,) ¿CHI4?
*CHI4:	tengo una pregunta: ¿cómo hicisteis las parejas? ¿por nivel de inglés?
*RES:	intenté poner gente que tuviera el mismo nivel, sí.
*CHI4:	pero tengo una pregunta: ¿quién tiene más nivel CHI6 o yo?
*RES:	de eso no me acuerdo no tengo aquí los… o sea la prueba esa que hicisteis antes de venir yo (,) ¿os acordáis?
*CHI4:	yo creo que tengo un poco más de nivel que CHI6.
*CHI5:	yo también creo que tiene un poco más de nivel.

*CHI4:	yo llevo yendo a la academia cuatro años.
*RES:	bueno da igual CHI6, también tiene un muy buen nivel (,) todos lo tenéis a ver (,) pues a ver los que vais a academia tenéis un poquito más eso es como siempre (,) ¿vale y ahora creéis que os ha ayudado en algo? ¿para vuestras clases de inglés?
*CHI2:	nos ha ayudado primero a trabajar en inglés, a saber más inglés…
*CHI1:	a expresarnos mejor.
*CHI6:	a mí me ha motivado más que nada.
*RES:	a ti motivado (,) más que nada ha sido motivación.
*CHI6:	sí.
*CHI4:	yo salvo la mayoría de las cosas ya las había hecho en la academia pero los *writings* se me dan bastante mal y pues para practicarlos me ha venido bien.
*RES:	vale muy bien, ¿muy bien alguna cosa más? ¿no? vale y ahora sobre el experimento en general os acordáis que aparte de esas dos tareas en parejas teníamos, por ejemplo, los ejercicios estos de gramatica John and Jenny, the snakes, luego también teníamos los *writings*, los cuestionarios de las caritas… entonces ¿qué os ha parecido todo eso en general?
*CHI3:	el de las caritas me pareció bastante.
*RES:	¿bastante? ¿demasiado pesado?
*CHI3:	sí, había bastantes preguntas.
*RES:	vale demasiadas preguntas (,) ¿algo más? ¿CHI4?
*CHI4:	es que se me olvidan las cosas que iba a decir.
*RES:	¿alguien más? ¿a alguien se le ocurre?
*CHI4:	¡ah sí! que ha venido bien que vengas porque el inglés en el colegio es muy bajo o no sé si nos dan las clases porque creen que es nuestro nivel o porque es lo que saben dar pero me parece que es muy bajo…
*RES:	tú te sientes que el nivel que exigen….
*CHI4:	es para niños de cinco años y a mucha gente le parece eso y no sé por qué no nos dan algo más, que somos niños de sexto mucho inglés, es muy bajo y esto a mucha gente le ha pillado de sopetón, porque es un poco más alto que este (,) no es superalto pero es más alto que este.
*CHI1:	yo cuando estaba en mi anterior colegio…
*CHI4:	es que aparte el inglés de clase es que… seguimos haciendo ejercicios de… eh… unir las… dibujos y colorearlos.
*CHI1:	en mi anterior colegio estaba…
*RES:	¿a cuál ibas?
*CHI1:	a XXXXX.
*RES:	aja.
*CHI1:	era como un nivel más alto y vine aquí y me parecía fácil más fácil que el anterior.
*RES:	¿entonces los ejercicios que habéis hecho conmigo se parecen a lo que hacéis en clase o no?
[they all answer no]
*CHI3:	más altos.
*CHI1:	muchísimo más alto el nivel que has dado.

*RES:	vale, y ya si para terminar: ¿cambiaríais algo de lo que hemos hecho? (,) si yo fuera a otro colegio a trabajar con otros niños me diríais: “Asier, pues cambia esto porque no nos ha gustado o esto ha sido…”
*CHI4:	en lo de las caritas [meaning the attitude questionnaire] al final del todo dejar una especie de textito para que la gente pueda poner como qué opina de alguna pregunta.
*RES:	ah vale (,) ¿sobre el cuestionario en sí? sobre las preguntas… eso es muy interesante (,) ¿algo más? ¿más o menos días? o hacerlo un poco más seguido?
*CHI2:	más seguido.
*CHI6:	yo creo que está bien.
*CHI1:	yo en lo de escuchar y escribir yo cambiaria en el primero unas parejas y en el segundo otras.
*CHI6:	para comparar si trabajan mejor o igual.
*RES:	vale, vale muy bien, muy bien, eso es interesante (,) CHI3 algo más que se te ocurra?
*CHI6:	¡ah yo sí! ¡más *chupachuses*!
*RES:	[laughs] ¿más *chupachuses*? ahora os voy a dar más *chupachuses* (,) ¿algo más? ¿cerramos aquí la sesión? vale.
@End

D12 Multiple regression analysis

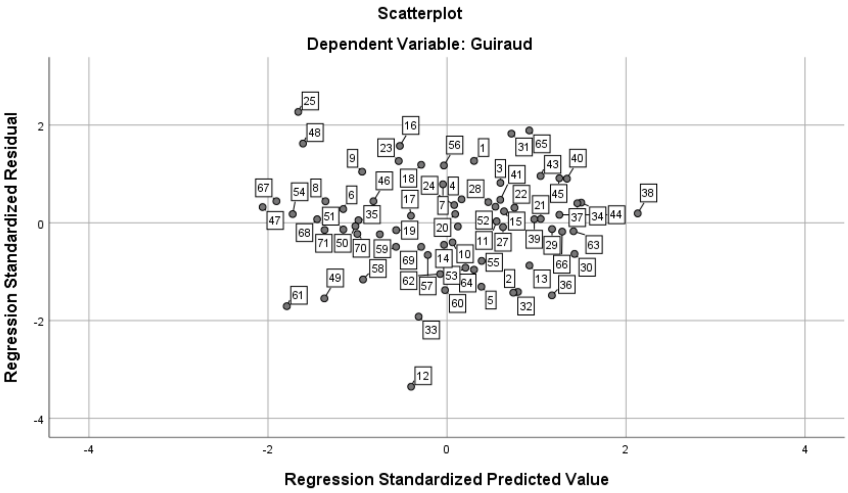
Multicollinearity check: VIF

Coefficients[a]

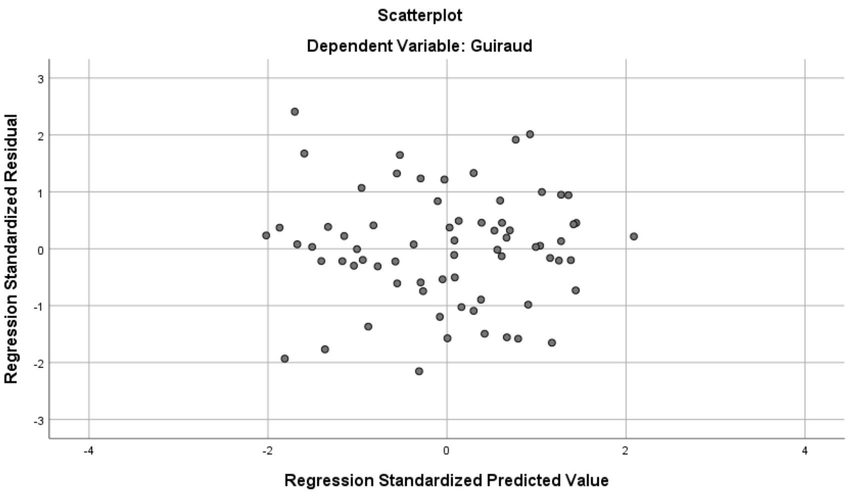
Model		Collinearity Statistics	
		Tolerance	VIF
1	WritingAtt	1.000	1.000

a. Dependent Variable: L1Writing

1st Regression analysis attempt: normality check



2nd Regression analysis attempt: normality check



Model Summary[b]

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Durbin-Watson
1	.515[a]	.266	.244	.775682970986	2.034

a. Predictors: (Constant). WritingAtt. L1Writing
b. Dependent Variable: Guiraud

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	14.585	2	7.293	12.120	.000[b]
	Residual	40.313	67	.602		
	Total	54.898	69			

a. Dependent Variable: Guiraud
b. Predictors: (Constant). WritingAtt. L1Writing

Coefficients[a]

		Unstandardized Coefficients		Standar. Coeff.			95.0% Confidence Interval for B		Correlations		
	Model	B	Std. Error	Beta	t	Sig.	Lower Bound	Upper Bound	Zero-order	Partial	Part
1	(Constant)	.977	.637		1.533	.130	-.295	2.248			
	L1Writing	.366	.153	.258	2.400	.019	.062	.671	.345	.281	.251
	WritingAtt	.626	.171	.393	3.661	.000	.285	.967	.450	.408	.383

a. Dependent Variable: Guiraud

Residuals Statistics[a]

	Minimum	Maximum	Mean	Std. Deviation	N
Predicted Value	3.13482403755	5.02356672287	4.06364932931	.459763190669	70
Residual	-1.67036426067	41.868019938469	.000000000000	.764358524165	70
Std. Predicted Value	-2.020	2.088	.000	1.000	70
Std. Residual	-2.153	2.408	.000	.985	70

a. Dependent Variable: Guiraud