Collaborative writing and feedback: A longitudinal study of the potential of models in primary EFL writing

María Luquin Urtasun

Supervised by:

Dr. María del Pilar García Mayo

Vitoria-Gasteiz 2022

Director: Dr. María del Pilar García Mayo
Universidad del País Vasco – Euskal Herriko Unibertsitatea (UPV/EHU)
Department of English, German and Translation and Interpretation

This doctoral dissertation has been funded by a scholarship from the Spanish Ministry of Economy and Competitiveness (Formación del Personal Investigador. Reference BES-2017-079672).

Acknowledgements

I would like to thank a number of people for their help and support during the production of this thesis.

First and foremost, I am extremely grateful to my supervisor, Professor María del Pilar García Mayo, for her valuable suggestions, priceless feedback and expert guidance and for helping me see the forest for the trees and make some sense of the confusion. But I especially thank Pilar for what I admire her the most, which is her ability to be a professional and a motherlike figure at the same time. Thank you for being my mentor, my guide, my consultant and my 'pilar', for pointing me in the right direction without demand, for your warm support, your sharing of stories and life experiences before, during and after our meetings, for teaching me to deal with one disaster at a time (I am still on it), and for having faith in me even in my darkest moments.

This dissertation is part of a larger research project entitled *Interaction and written production. The potential of collaborative writing in the learning of English as a Foreign Language by primary school students*, led by Professor García Mayo and funded by the Spanish Ministry of Economy and Competitiveness (FFI2016-74950-P). This project has as its main goal to explore the potential of collaborative writing in EFL learning and to document collaborative interaction among Primary School learners while they carry out tasks that lead to a written product. The present doctoral thesis has been possible thanks to the research grant I was awarded (BES-2017-079672) through which I became a member of the research group, thus providing me with the financial means to complete this dissertation.

The statistical analyses of this study could not have been completed without the help of Yacelli Bustamante, who patiently spent infinite hours with me trying to make this work somewhat 'significant'. Special thanks also go to Professor Vicente Nuñez Antón for teaching me that 'bipolar statistics' and a 'Cocktail Molotov' were nothing but 'bilateral statistics' and 'Kolmogorov-Smirnov' tests. Thank you for being always available to give some advice, to create some statistical knowledge out of this lettered mind, and to put me back on the right track when I was completely lost. I would also like to thank Dr. Kim McDonough for welcoming me at Concordia University, and for turning all my analyses upside down. Without your feedback, this work would have certainly been

easier, but way much more flawed. Thank you all three for guiding me through what has been the worst nightmare in this journey.

The empirical and longitudinal study work presented here required long hours of work at San Viator school in Vitoria-Gasteiz. I would like to thank all the teachers and pupils who have been involved in this project for their patience and involvement. In particular, I am very grateful to Ms. Eli Zubizarreta for always finding a space for research, no matter how busy she is. They all bore the brunt of this work. I also want to thank my assistants for helping me disinterestedly and just for the sake of knowledge. More people like them are indeed needed in this research world.

I would also like to acknowledge my colleagues and lecturers at the University of the Basque Country and the Public University of Navarre. Thanks for your words of encouragement, for those tips that helped furthering my research and for becoming good colleagues with whom I have been able to vent my feelings in my worst moments and who know what writing a thesis dissertation implies.

My family deserves a special mention. I am grateful for my parents' unconditional, undeniable, and loving support that kept me positive and confident, for their wise counsel and sympathetic ear, for being the main promoters of my dreams, for trusting and believing in my expectations, and for the values and principles that they have instilled in me. Thank you for teaching me to fight tenaciously to achieve my goals, for showing me that the desire to improve allows me to keep on going despite the obstacles that arise, for giving me the opportunity to have a complete training and for trusting me until the end. You are always there for me. Deepest thanks to my sister and brother, Gemma and Roberto, who keep me grounded, remind me of what is important in life, provide guidance and a sounding board when needed, and are always supportive of my adventures. Finally, I also owe my gratitude to my partner, Unai, my greatest discovery in these four years, thank you for constantly listening to me rant and talk things out during difficult times, for cracking jokes when things became too serious, for putting up with me and for helping me look on the bright side as I hurdled through the path of this dissertation and of life.

Last, but not least, I would like to thank those who are about to embark on this journey with me and to do your bit by reading and reviewing a jumble of letters trying to sort themselves to mean something one day. I just hope you bear it well…

TABLE OF CONTENTS

LIST OF ABBREVIATIONS

(In alphabetical order)

CAF = Complexity, accuracy and fluency

CALL = Computer-assisted language learning

CF = Corrective feedback

CG = Control group

CLIL = Content and language integrated learning

CRE = Content related episode

CW = Collaborative writing

LTG = Long-term treatment group

EC = Error correction

EFL = English as a foreign language

ESL = English as a second language

FL = Foreign language

FNs = Features noticed

FNIs = Features noticed and incorporated

FonF = Focus on form

FonFs = Focus on forms

FUI = Features unnoticed and incorporated

IL = Interlanguage

L1 = First language

L2 = Second language

LRE = Language-related episode

LTG = Long-term treatment group

MG = Model group

NNS = Non-native speaker

NP = Noun phrase

NS = Native speaker

PFNs = Problematic features noticed

TTR = Type Token Ratio

RTTR = Root Type Token Ratio

S1 = Stage 1

S2 = Stage 2

S3 = Stage 3

SCT = Sociocultural theory

SL = Second language

SLA = Second language acquisition

TG = Treatment group

TL = Target language

VP = Verb phrase

WCF = Written corrective feedback

YL = Young learner

ZPD = Zone of Proximal Development

LIST OF TABLES

LIST OF FIGURES

LIST OF GRAPHS

ABSTRACT

In the area of Second Language Acquisition, great attention has been paid to the possible potential of feedback in written production and its impact on the development of the learner's interlanguage (Polio, 2012; Van Beuningen, 2010). The argument is that feedback is related to the processes of attention to form, which is crucial for acquisition (R. Ellis, 2016). Recently, researchers have proposed that model texts, samples of well-written texts that learners compare with their original draft (Martínez Esteban & Roca de Larios, 2010), can be used as an effective written corrective feedback (WCF) technique. Actually, there are a handful of studies in which the effects of models have been analyzed in individual (e.g., García Mayo & Loidi Labandibar, 2017; Hanaoka, 2006a, 2006b, 2007; Kang, 2020; Montealegre Ramón, 2019) and collaborative (e.g., Coyle & Roca de Larios, 2014; Coyle, Cánovas Guirao & Roca de Larios, 2018; Luquin & García Mayo, 2021; Yang & Zhang, 2010) writing or both (Lázaro-Ibarrola, 2021, Martínez Esteban & Roca de Larios, 2010). Overall, the results have revealed that models are effective for directing attention to content, lexical features, and alternative expressions and ideas, and that collaboration furthers the learning potential of the model. However, studies on collaborative writing and models as a form of WCF have been conducted with adult participants, but research with children is scarce despite the growth of early EFL learning in school settings over the past twenty years. In addition, more work is needed to elucidate whether the benefits reported in these studies only lead to greater precision in L2 writing or to language development in the long term (Polio, 2012).

Besides the internal cognitive factors that intervene in the learner's attention and response to feedback, affective variables can also mediate in how students engage with corrections. In this sense, the notion of student commitment has received more and more attention from researchers who have analyzed cognitive responses and affective behaviors shown by students towards different types of feedback (Wigglesworth, & Storch, 2012; Tocalli-Beller & Swain, 2005). However, very few studies have explored children's motivation toward model texts (see Lázaro-Ibarrola & Villarreal (2021) and Villarreal and Lázaro-Ibarrola (submitted). Certainly, writing is a complex and time-consuming activity which requires concentration, even more so for YLs who are still developing cognitively, physically, and socially. Therefore, considering such factors when conducting research with children seems of paramount importance if we want to

draw a complete and more accurate picture of what is going on during the acquisition of an L2.

The research presented here represents an attempt to bridge these gaps in the field by examining the short- and long-term effects of collaborative writing and model texts on the L2 learning by young learners, measuring their motivation along the way.

The participants were 60 11- to 12-year-old Spanish children forming a total of 30 pairs from three EFL classes randomly assigned to a control group (CG), a treatment group (TG), and a long-term treatment group (LTG). The groups were engaged in two four-stage collaborative writing cycles of 3 weeks each separated by four months. The four-stage task involved (a) noticing of linguistic problems while writing a picture-based story (Stage 1), comparison of their texts with a native-speaker model (Stage 2)/self-correction of the students' own texts, (c) rewriting of their original output (Stage 3) and (d) delayed post-test (Stage 4). The CG did not receive the models, but self-corrected their own texts, the TG was only exposed to the feedback at two times (one per Cycle), and the LTG benefitted from this technique during the two writing cycles and the period in-between.

Given the lexical gains that the models seem to offer to the detriment of grammar, we also opted to guide the children towards formal aspects through the technique of input enhancement. We chose to work with third person singular possessive pronouns (*his/her*) since it is a grammatical structure that Spanish children find difficult to differentiate. In this way, the possessive pronouns present in all the models were underlined. With the aim to collect qualitative data on the children's attitudes and motivation toward the task, they completed an individual questionnaire before Stage 1 and after Stage 3 in both cycles. In addition, another individual questionnaire was administered and focus group interviews were held with six students from each group randomly selected after the two cycles were completed.

The results obtained show that the use of model texts brought about an increase in the number of language-related episodes (LREs) generated and a greater attention to lexical and content aspects. Nevertheless, after prolonged exposure, the models helped the children diversify their linguistic concerns and so it was reflected on their drafts, as many of the aspects noticed were fully or partially incorporated. On the other hand, self-

correction promoted attention to grammar, spelling and punctuation, which reflects the benefits of writing tasks to engage in *languaging*.

Regarding the effect of models on draft quality, in the short term the models led the children to reduce the number of pre-clauses and proto-clauses as well as to increase the grammatical complexity of their texts, which became visible in the high frequency of subordinate clauses. After sustained exposure to the models, the children were able to produce fewer proto-clauses and more clauses, have greater lexical diversity in their texts, and make fewer mistakes. With reference to input enhancement, no statistically significant differences were found between the texts written by the CG and the TG, but some trends were observed. However, we did observe statistically significant differences within the LTG as well as when the three groups were contrasted. Namely, the children in the LTG used all third-person possessive pronouns correctly in draft 6, reaching significance with respect to their first draft. Also in draft 6, they achieved better results than the CG and the TG and this result is likely to be attributable to the sustained exposure to the enhanced input.

Finally, concerning attitudes and motivation, the use of models elicited positive responses among the participants, especially in the case of the children who had been exposed to this form of feedback for a longer time. In general, some students showed a preference for a more explicit type of error correction, but their enjoyment, improvement in vocabulary, grammar, and overall written competence, along with their enthusiasm for working collaboratively, make models a very interesting tool to integrate into the EFL classroom. In light of these results, pedagogical implications will be discussed.

RESUMEN

En el área de Adquisición de la Segunda Lengua (ASL) se ha prestado gran atención al posible potencial de la retroalimentación en la producción escrita y su impacto en el desarrollo de la interlengua del aprendiz (Polio, 2012; Van Beuningen, 2010). El argumento es que la retroalimentación se relaciona con los procesos de atención a la forma, cruciales para la adquisición (R. Ellis, 2016). Más concretamente, la recepción de retroalimentación en forma de modelos está atrayendo la atención de muchos investigadores en los últimos años (Abe, 2008; Cánovas Guirao, 2017; Cánovas Guirao, Roca de Larios y Coyle, 2015; Coyle y Roca de Larios, 2014, García Mayo y Loidi Labandibar, 2017; Hanaoka, 2006a, 2006b, 2007; Hanaoka e Izumi, 2012; Martínez Esteban y Roca de Larios, 2010; Yang y Zhang, 2010). La razón de este interés se debe a la gran cantidad de ventajas que el texto modelo aporta en comparación con otras formas de corrección, limitadas a señalar o corregir el error del alumno sin tener en cuenta otros aspectos del espectro de una lengua (Hanaoka, 2007).

En este sentido, el modelo ofrece una alternativa al texto original del alumno, que después de haber hecho el esfuerzo de escribir en una lengua extranjera es capaz de comparar y analizar el modelo nativo. Este tipo de retroalimentación permite corregir errores y además ofrece alternativas a lo que ya es correcto en sus textos, llegando a ser un buen ejemplo de escritura nativa no solo a nivel oracional, sino también a nivel discursivo. En otras palabras, a través de los modelos, los alumnos reciben información sobre vocabulario, gramática, contenido, coherencia, cohesión, frases idiomáticas, estructuras oracionales, o simplemente están expuestos a otra forma de decir las cosas de una manera más nativa (Hanaoka 2006b, 2007; Yang y Zhang, 2010).

Recientemente, en el área de ASLs se está poniendo especial énfasis a planteamientos de tipo psicolingüístico que avalan el potencial del proceso de escritura para el aprendizaje de lenguas. Esta perspectiva se ha denominado ‘escribir para aprender lenguas’ (Manchón, 2011, 2014) en contraste con el estudio más tradicional de la competencia escrita en lenguas segundas/extranjeras que se resumiría en ‘aprender a escribir’ en esas lenguas. Manchón (2011) concluye que el proceso de escritura fomenta el procesamiento lingüístico que facilita el aprendizaje quizá de forma más clara que la producción oral del lenguaje. Por otra parte, se ha demostrado también que la ejecución de una tarea de forma colaborativa da paso al diálogo y a la negociación, y el mero

hecho de producir lenguaje y de hablar sobre el lenguaje hace que los alumnos consoliden su conocimiento sobre el mismo, formen hipótesis, las pongan a prueba y las acepten o rechacen según la retroalimentación que reciban sobre ellas (Cánovas Guirao, 2017; Storch 2002, 2005; Wigglesworth y Storch, 2012).

Con respecto a las diferencias individuales, también hay un interés creciente en cómo las actitudes de los estudiantes tienen un efecto sobre el producto final. La investigación ha demostrado que una actitud positiva hacia estos métodos de retroalimentación parece conducir a la incorporación de más elementos en las revisiones posteriores, mientras que un bajo nivel de motivación da paso a un rendimiento deficiente. (García Mayo y Loidi Labandibar, 2017; Hanaoka, 2007; Yang y Zhang, 2010). Además, García Mayo y Loidi Labandibar (2017) abordan la cuestión de si las actitudes negativas de sus participantes podrían adscribirse a su corta edad, en comparación con los altos niveles de motivación encontrados en los estudiantes de Hanaoka (2007) y Yang y Zhang (2010). Aun así, el hecho de que los participantes en los últimos estudios escribieran sus textos en colaboración puede haber influido mucho en sus actitudes positivas hacia la tarea en cuestión.

Finalmente, cabe mencionar que esta población merece especial interés puesto que los niños constituyen los cimientos sobre los cuales construir el andamiaje para el aprendizaje de lenguas (Copland y Garton, 2014). Además, en el contexto de inglés como lengua extranjera es casi inexistente la investigación que se centra en cómo los niños de cursos superiores de Educación Primaria (5º y 6º), cuando comienza su producción escrita en lengua extranjera, producen un texto escrito de forma colaborativa. Se necesita mucha más información sobre jóvenes aprendices puesto que cada vez hay un mayor número de colegios que introducen una lengua extranjera a edades tempranas (Enever, 2011; García Mayo, 2017; Muñoz, 2014; Nikolovy y Djigunovic, 2011). Como indica Butler (2015), este tipo de investigación será beneficiosa no solo para incrementar nuestro conocimiento teórico acerca de cómo adquieren los niños una lengua extranjera, sino que también tendrá implicaciones directas para las políticas educativas y las decisiones pedagógicas que se adopten para esa franja de edad. Como defienden Collins y Muñoz (2016), los estudios sobre jóvenes aprendices en contextos de adquisición de lenguas extranjeras en el aula deberían ser uno de los objetivos primordiales de investigación en las próximas décadas.

El estudio

Objetivos y preguntas de investigación

El principal objetivo de esta tesis es averiguar si los estudiantes de primaria mejoran su producción escrita a largo plazo a través de la escritura colaborativa y la recepción de retroalimentación en forma de textos modelo redactados por nativos. Asimismo, aprovechamos los textos modelo para analizar el impacto de la técnica conocida como 'realce del input' en la percepción y producción de pronombres posesivos de tercera persona del singular. Finalmente, el proyecto busca comprender cuál es y cómo evoluciona la actitud y la motivación del alumnado durante la realización de estas tareas. Para alcanzar estos objetivos, hemos formulado las siguientes preguntas de investigación:

Módulo 1: Conexión oral-escrita

1. ¿En qué aspectos lingüísticos se fijan los jóvenes alumnos de inglés como lengua extranjera cuando se enfrentan a una tarea sobre textos modelo?

   a) ¿Qué aspectos del lenguaje perciben cuando redactan un texto narrativo?
   b) ¿Qué perciben cuando comparan sus textos escritos con textos modelo o cuando se corrigen a sí mismos?
   c) ¿Existen diferencias entre ciclos en de cada grupo (grupo control, grupo tratamiento y grupo largo tratamiento) con respecto a la frecuencia, tipo y resolución de los episodios producidos durante la interacción oral?

2. ¿Cómo se relaciona el contenido percibido en las etapas de composición y comparación con sus textos revisados? (Impacto de la producción oral en la producción escrita)

Módulo 2: Los efectos de los modelos en la producción escrita de los niños

3. ¿Ayudan los modelos a mejorar la producción escrita de los estudiantes de inglés como lengua extranjera a corto y largo plazo?

4. ¿Juega la técnica del realce del input un papel en el hecho de que los niños perciban y mejoren el uso de los posesivos de tercera persona del singular?

Módulo 3: Actitud y motivación de los alumnos

5. ¿Cuál es la actitud de los niños hacia los tres tipos de retroalimentación? ¿La motivación de los alumnos cambia a lo largo del tratamiento?

Metodología

Participantes

Los participantes de este estudio fueron 60 niños bilingües (euskera y castellano) de 11 a 12 años (27 niños y 33 niñas) de tres clases de un colegio concertado ubicado en Vitoria-Gasteiz. Antes de empezar con el experimento, los estudiantes realizaron una prueba de nivel de inglés (Cambridge Young Learners English Assessment, 2018) de forma que se pudiese emparejar a cada alumno con otros de su mismo nivel, resultando en un total de 30 parejas. En base al resultado obtenido en la prueba de nivel (A2), los textos modelos fueron redactados o adaptados y posteriormente evaluados por los profesores.

Tarea y procedimiento

La secuencia típica de investigación con modelos se basa en una tarea de escritura consistente en tres etapas: (i) la percepción de problemas lingüísticos mientras se escribe una historia basada en imágenes (Etapa 1), (ii) la comparación de sus borradores iniciales con los textos modelo (Etapa 2), y (iii) la reescritura del texto original, que nos permite observar si los alumnos incorporan las soluciones ofrecidas por los modelos en los textos revisados (Etapa 3). Además de estas tres etapas, decidimos incluir una evaluación posterior (Etapa 4) dos semanas después de la Etapa 3 para dilucidar si los alumnos eran capaces de incorporar los conocimientos adquiridos en un contexto diferente, evitando así efectos de repetición de tarea, y para comprobar hasta qué punto el nuevo material aprendido se mantenía en un período corto de tiempo.

Con el fin de realizar un estudio longitudinal sobre modelos, esta secuencia se llevó a cabo en dos ciclos separados por cuatro meses: el Ciclo 1 tuvo lugar en enero y el Ciclo

2 en junio. Las cuatro etapas que componen cada ciclo se realizaron en tres semanas. A cada clase se le asignó aleatoriamente un tipo de retroalimentación: (i) el grupo tratamiento (GT) (n = 22) recibió modelos en la Etapa 2 de cada ciclo, es decir, en dos tiempos diferentes (enero y junio); (ii) el grupo de tratamiento largo (GTL) (n = 20) estuvo expuesto a los modelos durante un período de tiempo prolongado (de enero a junio); y (iii) el grupo control (CG) (n = 18) no se benefició de ningún tratamiento, sino que se limitó a autocorregir sus textos en la Etapa 2.

Ante las ganancias léxicas que parecen ofrecer los modelos en detrimento de la gramática, optamos por guiar a los niños hacia aspectos formales mediante la técnica del realce del input. Elegimos trabajar los pronombres posesivos de tercera persona del singular (*his*/*her*) puesto que es una estructura gramatical que a los niños españoles les resulta dificil diferenciar. De esta manera, los pronombres posesivos presentes en todos los modelos utilizados aparecían subrayados.

El mismo procedimiento que se utilizó para el primer ciclo se utilizó también para el segundo, con la diferencia de que se utilizaron diferentes imágenes y modelos. La duración de cada sesión fue de aproximadamente 15 minutos. Con el objetivo de recoger datos cualitativos sobre las actitudes y motivación de los niños hacia la tarea, estos completaron un cuestionario individual antes de comenzar la Etapa 1 y al finalizar la Etapa 3 en ambos ciclos. Además, una vez finalizado el Ciclo 2, se administró otro cuestionario individual y se realizaron entrevistas de grupos focales a seis estudiantes de cada grupo seleccionados aleatoriamente.

Resultados

Con respecto a la percepción de aspectos del lenguaje y de contenido (Pregunta de investigación 1), los resultados obtenidos demuestran que el uso de textos modelo trajo consigo una mayor producción de episodios lingüísticos y una mayor atención a aspectos léxicos y de contenido. No obstante, después de una exposición prolongada, los modelos también ayudaron a los niños a percibir aspectos formales del lenguaje y así quedó reflejado en sus textos, ya que muchos de los aspectos percibidos fueron parcial o totalmente incorporados. En consecuencia, parece que la capacidad limitada de procesamiento de los niños en estos niveles bajos de competencia del lenguaje, que parece conducir a una mayor atención al significado (VanPatten, 1990, 2004), podría

desbloquearse hasta el punto de ampliar su 'alcance de percepción' (Hanaoka , 2007) a aspectos gramaticales. Por otro lado, la autocorrección fomentó la atención a la gramática, la ortografía y la puntuación, lo cual refleja los beneficios de las tareas de escritura para percibir y trabajar aspectos lingüísticos.

En cuanto al impacto de la interacción oral en el ámbito escrito (Pregunta de investigación 2), la percepción de los aspectos lingüísticos y de contenido en las Etapas 1 y 2 resultó en una mejora de los borradores de los participantes, ya que muchos de estos aspectos se incorporaron total o parcialmente. Además, observamos que cuanto mayor era la exposición a los modelos, más se diversificaba la atención y más correctos eran los elementos incorporados.

Por lo que concierne al potencial de los modelos en la competencia escrita de los niños de Educación Primaria (Pregunta de Investigación 3), a corto plazo los modelos llevaron a los niños a reducir el número de pre-cláusulas y proto-cláusulas y a aumentar la complejidad gramatical de sus textos, lo cual se hizo visible en la mayor frecuencia de cláusulas subordinadas. Después de una larga exposición a los modelos, los niños consiguieron producir menos proto-cláusulas y más cláusulas, tener una mayor diversidad léxica en sus textos y cometer menos errores. Es decir, los niños se convirtieron en mejores escritores en inglés gracias a los modelos.

En relación a la técnica del realce del input (Pregunta de Investigación 4), no se encontraron diferencias estadísticamente significativas entre los textos redactados por el GC y el GT, pero se observaron algunas tendencias. El GC no solo no mejoró, sino que la precisión de sus textos empeoró a medida que se acercaba el final del experimento. El GT proporcionó un mayor número de respuestas correctas en el borrador 3, pero su rendimiento resultó ser más pobre en el Ciclo 2. Por el contrario, la percepción del GTL de los pronombres posesivos de tercera persona en el borrador 3 permaneció estable a lo largo de las composiciones restantes, llegando a un punto en el que los niños usaron todos los pronombres posesivos de tercera persona del singular correctamente. Es más, este resultado resultó ser significativamente diferente del obtenido en el primer borrador. Cuando se contrastaron los tres grupos, solo observamos diferencias estadísticamente significativas entre el GTL y los otros dos grupos en el borrador 6. Es decir, el GTL consiguió mejores resultados que el GC y el GT y es posible que este resultado sea atribuible a la exposición sostenida al input realzado.

Finalmente, en lo que respecta a las actitudes y la motivación de los alumnos hacia sus respectivos tratamientos (Pregunta de Investigación 5), el carácter aburrido y repetitivo de la autocorrección generó un sentimiento pesimista hacia la tarea, pero los alumnos reconocieron su eficacia para mejorar la ortografía y el vocabulario, entre otros aspectos. Por el contrario, el uso de modelos provocó respuestas positivas entre los participantes, más aún en el caso de los niños que se habían beneficiado de esta forma de retroalimentación durante más tiempo. En general, algunos estudiantes mostraron preferencia por un tipo más explícito de corrección de errores, pero su disfrute, la mejora en el léxico, la gramática y en su competencia escrita en general, junto con su entusiasmo por trabajar en colaboración, hacen de los modelos una herramienta muy interesante que podría integrarse en el aula de inglés como lengua extranjera.

Conclusiones

Con la presente tesis doctoral hemos intentado contribuir al conocimiento de tres esferas principales desde una perspectiva longitudinal: el potencial de los textos modelo en el aprendizaje de idiomas, la intervención centrada en la forma y las diferencias individuales.

El uso de modelos a corto plazo resultó útil para ayudar a los niños a percibir un gran número de problemas lingüísticos, ampliar su repertorio léxico e incorporar nuevas estructuras en sus borradores. Después de cuatro meses de exposición, observamos que los estudiantes diversificaron sus preocupaciones lingüísticas e integraron un espectro más amplio de características en comparación con el Ciclo 1 y también en contraste con otros grupos. En cualquier caso, teniendo en cuenta que nos enfrentamos a estudiantes que aún no han desarrollado al máximo sus capacidades cognitivas y que, por ende, necesitan procesar la retroalimentación durante una tarea centrada en el significado, sería recomendable guiar a los niños en el análisis de los modelos con el fin de hacerles conscientes de lo que los textos modelo pueden ofrecer. Por lo tanto, la instrucción guiada debería estar presente durante todo el proceso.

No obstante, los textos modelo no son una estrategia de retroalimentación personalizada hecha para adaptarse a las necesidades individuales. En este sentido es probable que, para los alumnos con más dificultades, esta técnica no siempre sea la solución más adecuada para corregir los errores de los niños, especialmente los errores gramaticales

(Cánovas Guirao et al., 2015). Es por ello que alentamos a los instructores a utilizar modelos en combinación con otras formas de retroalimentación directa e indirecta, focalizada y no focalizada, como la corrección de errores, reformulación o el uso de comentarios metalingüísticos, que puedan señalar más específicamente (y ofrecer soluciones a) problemas gramaticales. De esta forma, conocer los beneficios que ofrecen las diferentes formas de retroalimentación y saber qué técnica es la más adecuada para las necesidades de nuestros alumnos debe ser una prioridad en la agenda de los profesores de idiomas.

El mero hecho de escribir también tiene efectos beneficiosos para el aprendizaje de una segunda lengua, y no es menos cierto en el caso de la autocorrección. Aunque la mayoría de los niños expresaron su descontento hacia esta práctica, es importante poner de manifiesto el potencial que tiene la autocorrección para el aprendizaje de una lengua extranjera. Incluso viéndose privados de orientación y debiendo afrontar la corrección solos, los niños fueron capaces de prestar atención y de solventar problemas en su mayoría gramaticales y ortográficos, lo cual indica que esta tarea es útil para trabajar el lenguaje y dirigir la atención de los niños hacia elementos formales.

Al hilo de este tema, uno de los desafíos que se presenta en el aula y en el campo de la investigación es cómo conectar la forma gramatical al significado durante tareas fundamentalmente comunicativas. En este punto sabemos que los textos modelo no son particularmente efectivos para dirigir la atención de los estudiantes hacia la forma, al menos a corto plazo, pero sí que han demostrado ser un escenario ideal para utilizar la técnica del realce del input, mejorando así la percepción y el uso de construcciones lingüísticas específicas a largo plazo. En particular, el aumento de la prominencia visual de los posesivos de tercera persona en los textos modelo ocasionó una mayor atención a dicha construcción. Los resultados, por lo tanto, indican que el realce del input, cuando se aplica bajo ciertas condiciones (alta prominencia de la estructura, conocimiento previo, subrayado, exposición continua, etc.), puede ser una intervención eficaz de atención a la forma. Realzar construcciones gramaticales específicas en textos modelo puede dirigir la atención de los niños hacia estas características y, por lo tanto, fomentar el desarrollo de una segunda lengua con el tiempo. Por consiguiente, basándonos en los resultados del presente estudio, sería aconsejable que los profesores de inglés como lengua extranjera hicieran uso de una instrucción sostenida implícita y explícita centrada en la forma. De esta manera, pueden brindar a los niños la ayuda suficiente

para lograr aquellos objetivos que puedan presentar más dificultades debido a sus propias limitaciones cognitivas (o incluso a las de la retroalimentación).

Finalmente, el procesamiento de la retroalimentación por parte de los estudiantes está influenciado por factores relacionados con el contexto y con el propio individuo. Por lo tanto, la consideración de factores individuales y contextuales constituye un intento interesante de lograr una comprensión más integral de la relación entre la retroalimentación y el aprendizaje de una segunda lengua, más aún dentro del contexto de los niños. En esta línea, Shak (2006) destacó que la educación infantil no se puede concebir sin la idea de motivación. De hecho, si extrapolamos este fenómeno a nuestro estudio, encontramos las dos caras de la moneda. Es decir, por un lado, observamos que la autocorrección generó una actitud pesimista en nuestros alumnos. Ante este rechazo, proponemos que los profesores transformen la autocorrección en una tarea más divertida y productiva como la corrección entre compañeros, y que eviten hacer un uso extensivo de esta técnica de retroalimentación. Por otro lado, el uso de modelos trajo consigo altos niveles de motivación, más aún en el caso de aquellos niños que estuvieron expuestos a esta técnica durante un período prolongado de tiempo. Así pues, es importante que los niños se sientan motivados durante el aprendizaje, de tal manera que puedan desarrollar un nivel sostenido de motivación necesario para un mayor rendimiento académico (Bitchener & Storch, 2016). Es responsabilidad pues del profesor llevar a la clase tareas que los jóvenes encuentren útiles y atractivas. En lo que respecta a la escritura, esto puede entrañar mayor dificultad puesto que los niños no suelen responder tan bien a las actividades escritas como a las orales dada la falta de dinamismo de aquellas, pero cada vez estamos más cerca de hacer que las tareas escritas sean más atractivas y efectivas.

Con todo, esperamos que nuestros hallazgos sean útiles no solo para los docentes e investigadores, sino también para las autoridades educativas que necesitan tomar decisiones sobre las mejores prácticas pedagógicas en el aprendizaje de lenguas extranjeras.

INTRODUCTION

The worldwide introduction of a foreign language (FL) at an early age responds both to a very clear social demand and to the conviction that young children have greater auditory and oral plasticity, which will allow them to assimilate the FL better than at a more advanced age (Nicholas & Lightbown, 2008). It is true that young learners (YLs) enjoy great facility for the development of basic communication skills, such as oral comprehension and production, fundamentally linked to social skills (Kellogg, 2008). As a matter of fact, in the Spanish educational context, among the four skills that are usually trained throughout the academic life, the writing skill has been neglected for the past few years in favor, almost exclusively, of the speaking skill (Cánovas Guirao, 2017). Nowadays, there seems to be a growing interest in the teaching of English writing due, in large part, to the recognition of the positive influence that its practice has on the overall development of language learning/teaching (Sharwood-Smith, 1974). Nevertheless, many English as a Foreign Language (EFL) teachers still appear to ignore that writing practice and written corrective feedback (WCF) play a fundamental role in the linguistic development of their students. Consequently, this practice does not have the room it deserves in current pedagogical agendas.

The lack of importance given to the writing skill in the EFL classroom comes into conflict with modern research in the field of second language acquisition (SLA). Oriented toward a writing-to-learn perspective (Manchón, 2011), the cognitive strand of SLA research is based on the assumption that learners need to receive information about the accuracy and appropriateness of their texts in order to drive their L2 knowledge forward (Ferris, 2010). One way of achieving this is through feedback provision, which seems to activate cognitive processes such as hypothesis formation and testing, attention, metalinguistic reflection, noticing, or problem-solving strategies (Williams, 2012). The act of trying out their hypotheses makes learners aware of what they actually know and do not know about the target language (TL), this way raising their awareness of ‘holes’ in their interlanguage (IL) and ‘gaps’ between their IL and the TL, a necessary process for acquisition (Schmidt, 2001). Due to the potential it represents for the promotion of language acquisition, WCF has been a widely studied subject in recent

decades (Bitchener & Knoch, 2009). It is from this assumption that the contribution of WCF to the language learning of children would seem a fruitful issue to explore.

As a complement to the cognitive perspective, the socio-cultural perspective, based on the work by Vygotsky (1978), emphasizes the social nature of the interaction and the co-construction of the learning process. Swain (2000) describes the process of solving problems and constructing knowledge through collaboration as learning in itself. It is precisely when the students share the responsibility of the oral or written product of a task that they collaborate in the solution of the problems that are presented to them, thus producing jointly discussed output that enables learners to solve problems and test hypotheses on which they receive feedback (Storch, 2016). The mutual help provided (also known as *scaffolding*) allows the linguistic knowledge to be extended beyond the level that each one of the learners has. Writing entails a complex, meaning-making process in which a range of social and cognitive elements play a part (Byrnes & Manchón, 2014). Within the context of YLs, children are still developing cognitively and even writing in their first language (L1) takes them considerable effort. Actually, research has shown that collaboration in L1 instructional contexts is beneficial for children's writing since it fosters learners' engagement with the task, reduces the cognitive burden, reinforces affective variables such as motivation and leads to upgraded texts (Yarrow & Topping, 2001). In English as a second language (ESL) and EFL contexts, studies about collaborative writing with adults and adolescents (e.g., Santos, López Serrano & Manchón, 2010; Storch & Wigglesworth 2010a; Swain & Lapkin, 2002; Wigglesworth & Storch, 2012) have revealed how the sharing of knowledge and combined reflection on language use, or what Swain (2006) has termed *languaging*, during writing and feedback processing, might relieve the otherwise complex processing demands made on individual writers. However, we know very little about the effects of children's oral interaction on their joint written product (e.g., Cánovas Guirao, Roca de Larios & Coyle, 2015; Coyle, Cánovas Guirao & Roca de Larios, 2018).

There are different ways in which feedback on L2 learners' writing can be provided. Methodologies of WCF may vary, for instance, as regards their focus, and their degree of explicitness including direct, indirect and metalinguistic feedback types. Research on feedback has posed a series of limitations as to its effectiveness when it is delivered in

the form of correction of errors. Although positive results have been reported regarding the effects of this feedback on a limited number of linguistic forms, some researchers have found fault with the merits of error correction (EC). Consequently, the value of traditional EC for L2 development is still a vexed question (Martínez Esteban & Roca de Larios, 2010; Polio & Williams, 2009).

As mentioned above, given that students need to notice the gap between their IL and the TL and test their hypotheses, along with the fact that the relationship between WCF and L2 development remains an open question, other feedback techniques should be deployed at least as partial options to traditional feedback methods. One of the alternatives that has aroused special interest among researchers are model texts (e.g., Hanaoka, 2006a, 2006b, 2007; Hanaoka & Izumi, 2012), since there is evidence that they play an instrumental role in promoting noticing and metalinguistic awareness, and in engaging students in deeper processing than with more traditional feedback methods such as EC (Hanaoka, 2007a; Hanaoka & Izumi, 2012; Martínez Esteban & Roca de Larios, 2010; Qi & Lapkin, 2001). This more discursive feedback method does not explicitly single out errors but treats the text as a whole providing appropriate language, organization, mechanics, style and ideas for a given context, rather than offering lists of corrected errors, editing symbols or metalinguistic codes (Cánovas Guirao, 2017).

In addition, the potential effect of feedback on L2 development has been evidenced to be arbitrated by intervening learner internal and external factors that affect how learners process feedback, the extent to which they notice gaps and their capacity to exploit the opportunities made available to them. Studies about children's attitudes to WCF are scant, so further research in this field would be of great help in determining whether affective factors, such as motivation, will have an effect on their performance and, ultimately, on their learning process (García Mayo, 2018; Lázaro-Ibarrola & Villarreal, 2021). What is more, we also find this call within the sociocultural perspective, when researchers criticize the studies on WCF for being decontextualized, and for not providing information about the context, teachers or learners' goals and beliefs that may ultimately exert an influence on the provision of and response to WCF.

Finally, firm causal evidence on the long-term positive impacts of writing practice and feedback is yet unavailable (Manchón, 2011). The results of previous studies using only

one-shot feedback treatment can only be interpreted as evidence of learners' uptake rather than acquisition (Reinders, 2009). Therefore, in order to analyze feedback for acquisition, we would need a study of children's performance over a longer period of time that would enable the tracking of changes in their written production. Following Bitchener and Storch (2016) and Ortega (2012), this progress may be tracked down in the inconsistent, nonlinear and irregular use of the new L2 forms by the children. In doing so, the present study aims to contribute to increasing our knowledge about the effects of models on YLs and bring to light an uncharted research niche regarding the lasting effects of model revision.

To sum up, we have brought to the fore some issues of the SLA domain of which we still know very little, such as the impact that child oral interaction has on their written product or how primary school-aged children write in collaboration and benefit from WCF in the EFL context. There is also a need to explore alternative WCF techniques, such as models, to shed light on the inconclusive findings on EC, to analyze the benefits of those techniques for YLs and to examine the characteristics of their interaction when dealing with feedback of this type. Additionally, in order to consider any results as evidence of acquisition and not only of uptake, we have highlighted the researchers' calls for longitudinal studies, that would enable the tracking of changes in the YL's written production. All in all, the questions raised above need to be addressed to fully comprehend the paramount role that collaborative writing and model texts play in the language learning experience of child EFL learners.

With these research gaps in mind, the present study aims to investigate the lasting benefits of modelling on children's L2 development. In particular, this dissertation wants to (i) explore the impact that child oral interaction has on their written product; (ii) obtain evidence on whether long-term engagement with writing practice and feedback can bring about learning; and (iii) analyze the effect of input enhancement on the students' use of third person singular possessive pronouns. The present study will also attempt to explore the role of pedagogical intervention such as pre-task instruction, input enhancement or collaborative reflection to promote the noticing and subsequent incorporation of formal features into the children's drafts. Finally, in view of the scarcity of studies about children's attitudes to WCF along with the fact that motivation seems to be fundamental for the students' consolidation of their new knowledge

(Bitchener & Storch, 2016), we will also collect our learners' beliefs and opinions on the specific tasks they performed in order to gain more insights into the motivational processes occurring in the L2 classroom.

In order to do so, three groups of 60 EFL Spanish learners worked in proficiency-matched pairs in two four-stage writing cycles over the span of six months. During these 4x2 stages, we explored the merits of model texts through the analysis of the production, nature and outcome of the LREs generated in oral interaction; the features incorporated into subsequent writings; the overall draft quality; and the noticing and use of third person singular possessive pronouns. Finally, in order to obtain qualitative data on certain individual differences, the children's attitudes towards the feedback and motivational disposition were measured along the way. All the data were submitted to the corresponding statistical analyses.

The main findings of this study reveal that models, in combination with collaborative work, have made it possible to improve many aspects of the children's developing L2 in terms of noticing, writing, learning and incorporation of new L2 features, consolidation of preliminary intake, focus on form, etc. In addition, upon looking at the children's performance, it is encouraging to observe that the participants improved their use of third person possessives as well as other formal features and expanded their vocabulary repertoire. What is more, although performance was not as optimal as that of adult learners given the YLs' limited abilities, we have obtained satisfactory results proving that a long exposure to model texts may be effective in alleviating children's constraints when processing feedback. The findings also showed that, in general, the participants were positive about the collaborative writing experience with models. This was not only a consequence of both dealing with models and working in pairs, but also a reason for taking out the most of this feedback technique. The analysis of motivation undertaken here has extended our knowledge of models, strengthened the value of collaborative writing and spurred on the implementation of motivation measures in task-based research. All these findings will be explained in detail in the following chapters.

The current thesis is organized as follows: in Part I (*Literature Background*), Chapters 1, 2, 3 and 4 provide the background for the four main issues of relevance to our study. Chapter 1 analyzes the importance of conducting studies on children's L2 acquisition.

In particular, we highlight that much more information is needed about YL as more and more schools are introducing a FL at an ever-earlier age. There is a relatively small body of literature that is concerned with what the developing child actually does when faced with specific tasks that facilitate the learning process. In consequence, we support the idea that studies on YLs in contexts of FL acquisition in the classroom should be one of the primary research objectives in the coming decades. Chapter 2 offers a theoretical overview of the role of WCF in L2 learning. This theoretical perspective rests on both cognitive and sociocultural constructs on learning such as the Interaction Hypothesis, input, output, noticing, Skill Learning Theory, Sociocultural Theory or collaborative writing. This section ends up bringing together the theories outlined in the chapter in order to explain how the interplay between these constructs may facilitate L2 learning. Chapter 3 presents a more practical perspective of WCF by zooming in on the nature of the different types of WCF that can be found in the literature. We also address the limitations of traditional feedback techniques and explain the reported benefits of modeling as an alternative, and more discursive feedback technique than EC. The chapter concludes with an overview of current research on models and highlights some shortcomings that this dissertation aims to address. Finally, in chapter 4 we explore the importance of individual differences in L2 writing. In particular, we describe how learner-internal factors can impact on the students' response to WCF. Within these internal factors, we examine cognitive variables such as working memory, developmental readiness or their level of L2 proficiency. Cognitive factors, however, also interact with affective variables such as motivation, which seems to be crucial for the students' consolidation of their new knowledge.

Part II, *The present study*, is comprised of three chapters: Chapter 5, *The study*, describes the methodology we have followed. First, we describe the rationale for the present research. Afterwards, the main goals and research questions will be posited, followed by the hypotheses entertained. Then the study itself is introduced: participants, materials, procedure and explanations about data analyses and codification. Chapter 6, *Results*, presents the results obtained to answer our research questions, which will be discussed in Chapter 7, *Discussion of main findings*, in relation to the Hypotheses posited. Finally, Part III (*Conclusions and contributions*), consists of one chapter only: Chapter 8, *Conclusions*, which provides the final conclusions in this dissertation and

acknowledges its limitations. Future directions for research on YLs and pedagogical implications for teachers and policymakers will also be suggested.

PART I

LITERATURE BACKGROUND

CHAPTER 1: CHILD L2 ACQUISITION

From its emergence in the 70s, child SLA has been overwhelmingly outshone by adult and adolescent SLA, becoming something like a 'Cinderella' area of study (Copland & Garton, 2014). However, given the growing number of schools introducing foreign languages (FLs) at an early age (Cameron, 2003; Enever, 2011, 2018; García Mayo & García Lecumberri, 2003), it comes as a surprise that only in the last few years has the research lens been placed on young learners (YLs) (Enever & Lindgren, 2017; García Mayo, 2017; 2018; Murphy, 2014; Pinter, 2005, 2007, 2011, 2014) in this context. As indicated by Collins and Muñoz (2016), studies on young FL learners in the classroom setting should be one of the primary objectives of research in the coming decades. This view is also underpinned by Oliver and Azkarai (2017, p. 1) who wrote that 'within the field of second language acquisition […], there has been much less research undertaken with children than with adults, yet the two cohorts are quite distinct in characteristics and in their learning processes'. Consequently, child second language (L2) learning deserves special attention in its own right. In this first chapter, we will explain the rationale for the interest in research on children and provide a brief description of the main traits of this age group as well as the differences between young and adult learners. We will conclude the section by summarizing the major findings of recent studies undertaken so far on child L2 acquisition in FL contexts.

1.1. The rationale behind early language learning

The amount of FL programs designed for children, predominantly with English as the target language (TL), is increasingly proliferating all over the world (Enever, 2018; García Mayo, 2018; Pinter, 2011). Furthermore, government policy worldwide stands up for an introduction of English into primary or even preschool at an ever-earlier age (Copland & Garton, 2014; Muñoz, 2014). The reasons for this 'hastiness' have been put forward in the literature. One of the strongest arguments is the one provided by governments and, as a result, by parents and society that learners need to be competent in the FL so as to have access to an international education and employment opportunities in this globalized world (Copland & Garton, 2014; Enever & Moon, 2009; García Mayo, 2018). A second reason is the belief that younger is better, grounded on

the significative benefits found in immersion and bilingual contexts (Lyster, 2007). Nevertheless, there is no conclusive evidence for the potential benefits of learning a FL at an early age (Copland & Garton, 2014; DeKeyser, 2013), since the large number of students per classroom, the limited access to the TL and the reduced opportunities to interact outside the classroom, or the scarce curriculum time available are some of the factors that hinder language learning (García Mayo, 2018; García Mayo & Hidalgo, 2017; Huang, 2015). In addition, age is also a variable that has proved to play a significant role in child SLA research, since it has been attested that the robust finding of 'the younger, the better' does not apply to all situations (García Mayo & García Lecumberri, 2003; Muñoz, 2006). Therefore, venturing into extrapolating these results to FL contexts may not be so tenable.

Notwithstanding the above, we cannot ignore the potential benefits derived from the early introduction of FLs in primary schools, such as the exposure to different cultures children can learn from and connect with, the positive attitudes and tolerance they can build toward languages (García Mayo, 2018; Nikolov, 1999) as well as the development of language strategies (Kearney & Ahn, 2014). Besides, children never come to the classroom empty-handed. They bring with them well-established skills or capacities which help them decode the FL, such as the ability to grasp meaning by drawing on intonation, gestures, actions and facial expressions (Halliwell, 1992), or their disposition to concentrate on meaning rather than accuracy (García Mayo, 2018).

All things considered, the assumption drawn from bilingual or immersion contexts that YLs are like sponges and will soak up the FL does not hold when transferred to settings where learners have limited exposure to input and this input is provided in large group contexts (Copland & Garton, 2014). Introducing a FL into the primary curriculum may pose benefits for children eventually, but there is hardly any research about what children really do when they face specific tasks that facilitate the learning process, and we also need to know more about effective pedagogies for teaching YLs in such restricted contexts and taking into account their specific needs as compared to adults (García Mayo, 2018). To that end, research on child FL learning becomes key to informing policymakers and to guaranteeing programs for children in FL contexts a place in the SLA field. YL's distinctive characteristics as well as the major differences identified between adults and children are described in the following pages.

1.2. What is a young learner?

As G. Ellis (2014) points out, the term 'young learner' is quite imprecise and can lead to confusion because, although it is true that a YL is legally anyone under the age of 18, using this term to cover such a wide and diverse age range is useless and impractical. Rather, G. Ellis suggests that a distinction between pre- and post- 11- or 12-year-olds should be made. The reason behind this boundary is the great disparity in 'physical, psychological, social, emotional, conceptual and cognitive development' (G. Ellis, 2014, p. 75) found among children in these age groups, resulting in very different approaches to teaching (Copland & Garton, 2014).

On the other hand, the context or even the place where this population is considered is also a factor to determine the age range within which learners are regarded as 'young'. For example, Europe, where this study is situated, agreed that pre-school children aged between three to six years old would fall into the 'very young learners' category, while primary school students aged between seven to twelve would be labelled as 'young learners' (Nikolov & Mihaljevic-Djigunović, 2011). Notwithstanding this distinction, in some contexts 14-year-olds are also categorized as being 'young learners'. Actually, in their review of child instructed SLA, Oliver, Nguyen and Sato (2017) decide to extend the threshold around the beginning of secondary school (14 years) under the premise that the decline of ultimate attainment in SLA depends on the age of onset (i.e., the age when acquisition begins), this decline being gradual, and not occurring at a specific age. What is more, research has shown that the advantages YLs may have tend to disappear by the age of 16 (Pinter, 2017).

In the same way as the European member states, G. Ellis (2014, p. 77) draws a distinction between 'early years/pre-primary' and 'primary' and lists a number of reasons for this. To start with, G. Ellis observes that younger children are the cornerstone of the 'younger is better' debate, which argues that at an early age, the child will be able acquire a FL with the same ease as they learn their mother tongue. Second, most studies until now have centered on secondary-school students and there is a general lack of information concerning the acquisition of a FL on the part of pre- and primary school students. Furthermore, the vertiginous lowering of the age at which English is brought to children worldwide, along with the spread of preschool English,

indicate that it is timely to concentrate on younger learners. Last, because of the different pedagogical approaches required to teach pre-, primary- and secondary-school students, concentrating on only one of these age groups enables a more exhaustive look at certain issues, such as suitable pedagogies and contexts of learning and avoids over-generalizations. (G. Ellis, 2014). Echoing the words of Enever and Moon (2010, p. 2), 'more precise descriptors are needed today, to ensure that age-appropriate approaches to teaching and learning are fully developed'. Therefore, for the purpose of this doctoral thesis, YLs will be considered those at pre-school and primary school level, roughly from the age of 3 up to 12 years old. Having defined what a YL is, let us now turn to consider what the main differential traits between young and older learners are.

1.3. Differences between children and adults

The area of SLA presents particular difficulties for those working with YLs in the sense that the L2 child's language, emotional development and socio-cognitive behavior are not as entrenched as that of an L2 adult, leading to significant individual linguistic variability (Oliver et al., 2017; Simon, 2010). Along similar lines, Nicholas and Lightbown (2008, p. 46) point out that 'for young learners, language acquisition involves cognitive, social, and physical engagement over long periods during which many changes take place in the developing child'. That is, childhood is a period of numerous and constant cognitive and physical changes. Therefore, while some tasks may be of help to a particular age group, others will not be very facilitative of their language development (Muñoz, 2007; Nicholas & Lightbown, 2008; Pinter, 2006).

By way of illustration, writing, which is the core issue of the present study, is said to entail a complex, meaning-making process in which a range of social and cognitive elements play a part (Byrnes & Manchón, 2014). Due to the continual process of cognitive development, YLs' first language (L1) writing skills are still developing throughout the adolescence period (Kellogg, 2008; Michel, Kormos, Brunfaut & Ratajczak, 2019). Through instruction at school, children also become more proficient in writing as a technical skill, and in text composition (Michel et al., 2019). Thus, this cognitive development along with the expertise gained through the years are believed to

have an effect on syntax, lexis and discourse aspects (Kellogg, 2008; McCutchen, 2011). For this reason, writing in a FL at this young age involves a cognitively demanding task, so we cannot deploy the same pedagogical approaches with children as with adults (or even adolescents) and expect the same results.

Besides the cognitive development, older and younger learners also differ in their language aptitudes and their learning strategies. While the former draw more on their analytical capacity to learn a language, the latter tend to resort to their memory (e.g., use of language chunks) (Michel et al., 2019). This issue constitutes one more reason to think that, for pedagogical approaches to be successful, tasks in the TL should be age-specific, and assessment formats and procedure should be also adequate for the corresponding YLs' cognitive and developmental stage (Butler & Zeng, 2014, Muñoz, 2007).

Another major aspect that has an effect on child L2 acquisition is the socio-cultural context, which varies depending on the age group (younger learners, older young learners and adults) and determines the way students interact with their interlocutors (Philp, Oliver & Mackey, 2008). Diverse internal and external factors, including the schooling experiences that go along with maturation, influence the way learners interact, and as a result, the way they acquire a language (Nicholas & Lightbown, 2008). Some other factors are the interlocutors' age and status (e.g., teacher, peer, parents), as well as the student's degree of independence in their interactions (Hidalgo, 2017).

The different cognitive developmental stages and the changes that children experiment in the socio-cultural context go hand in hand and, as mentioned above, these changes are visible in their interactions (Philp et al., 2008). At the same time the children are developing their L1, they are also learning how to interact with other people. Actually, if we observe how children interact, we will realize that they are less restricted by social conventions than their adult counterparts and, therefore, by task conditions (Nicholas & Lightbown, 2008; Philp et al. 2008). For instance, children are more likely to openly disagree with their partners, change the topic of the conversation, or even try to cheat when performing a task (Hidalgo, 2017). When research has compared child learners to adult learners on this matter, results indicate that, although YLs are capable of engaging in conversation cooperatively and use all the negotiation strategies identified in adult

studies, they do so to a lesser extent (Oliver, 1998; Pinter, 2006). In addition, Pinter (2006) observes that these two age groups understand and complete tasks in different ways: whilst adults seem to be more efficient, children do not follow an apparent order and use less time and language.

As we can see, child SLA differs significantly from adult SLA in several aspects, since the acquisition of the L2 happens concurrently with the child's cognitive, psychological, linguistic and social development. Apart from these factors, we cannot overlook either the fact there exist individual differences for adults and children, and these differences (attitudes, motivation, learning strategies, aptitude, anxiety, etc.) also need to be kept in mind (see Chapter 4). Therefore, when it comes to teaching and designing new materials, the particular needs, capabilities and perspectives of this young population need to be considered. (Muñoz, 2007, 2014; Nicholas & Lightbown, 2008; Pinter, 2011). In the section below, we will look at how some researchers have approached the study of child L2 acquisition in comparison with adult L2 acquisition, that way contributing to improving pedagogical practices both for children and educators.

1.4. Research on child L2 acquisition in FL contexts

Until recently, and notwithstanding the acknowledged differences between adult and child learners, research on child FL learners had been an uncharted territory in the SLA field (Philp, Oliver & Mackey, 2006). Most results derived from studies involving adult learners or high school students have been applied as the grounds for teaching methodologies, while studies pertaining to children in the FL context remain comparatively insufficient (Gagné & Parks, 2013; García Mayo & Lázaro-Ibarrola, 2015; Mackey & Oliver, 2002; Pinter, 2007). Also, on the pedagogical front, there is a necessity to look into how best to teach YLs and which conditions are facilitative of their learning. Teachers need to understand SLA theories and children's social and cognitive development in order to teach successfully, since a four-year-old child will differ from an eight-year-old one and the approaches that are valid for the latter will not be so for the former (Copland & Garton, 2014). Consequently, it is not viable to transfer the results from the extensive bulk of adult research and employ them as the cornerstone

for child directed instruction (Copland & Garton, 2014; Oliver, 2002; Oliver & Azkarai, 2017). Let us now consider the exceptions to this issue and zoom in on some of the major topics dealing with child L2 acquisition as well as those in need of further research.

From Long's (1983, 1996) work onwards, numerous studies have acknowledged that conversational interaction and communication facilitate L2 development (García Mayo & Alcón Soler, 2013; Loewen & Sato, 2018; Long, 1996; Mackey, 2007, 2020; Mackey, Abbuhl & Gass, 2012; Pica, 2013). Interaction allows the learner to map form to meaning, to manipulate their own production (modified output) to overcome communication breakdowns through negotiation of meaning, and to receive feedback on their communicative attempts in the TL. These three factors (input, output and feedback) are essential elements in the SLA process (R. Ellis, 1994; Oliver, 1998, 2009; Oliver & Mackey, 2003; Pica, 2013). As García Mayo (2018) observes, in FL contexts, where exposure to input is low, it is essential to supply YLs with as many learning opportunities as possible and interaction certainly plays a significant role in such process. Until relatively recently, the vast majority of studies on interaction have been conducted with adult participants and their results have been applied to children and used to inform pedagogy. However, one line of research to arise relatively early in child SLA was that revolving around the Interaction Hypothesis (Oliver et al., 2017). Among the most important studies we find the pioneering research carried out by Oliver and colleagues in the Australian English as a second language (ESL) context (Oliver, 1995; 1998; 2002, 2009), the work done by García Mayo and colleagues in English-as-a-foreign-language (EFL) contexts (see García Mayo, 2018 for a summary) or the research done in Canadian immersion programs (Lyster, 2007).

One of the first studies on children negotiating to perform a task was that of Van den Branden (1997), who looked into the effects of several types of negotiation on learner's output. The work focused on three groups of 11-12-year-old learners of Dutch who were asked to describe a series of pictures to a partner. Results showed that the participants who had been pushed in previous negotiations produced a significantly higher amount of output, made use of a greater range of vocabulary and were more prone to recycling the new language they had picked up than those learners in a comparison group who had not been pushed. However, no significant improvements were found on the

complexity nor the grammaticality of the learners' output, since the children did not seem to focus on language form.

In a series of studies carried out by Oliver and colleagues in the Australian ESL context children's ability to negotiate for meaning and benefit from feedback was proved to be different from that reported for adults, who used strategies in different proportions. For example, a study by Oliver (1998), who compared the negotiation by 196 L2 children aged 8 to 13 years with that by L2 adults reported in Long (1983), revealed that the YLs used fewer clarification requests and confirmation checks. In a follow-up study, Oliver (2002) examined conversational interactions among the same participants, paired to form 96 dyads of NS and NNS speakers (32 NNS-NS, 48 NNS-NNS, and 16 NS-NS dyads). The findings suggested that NNS-NNS pairs were more predisposed to engaging in negotiation for meaning, modifying their output to solve their breakdowns in communication.

Oliver (2000) analyzed whether there were any differences in the provision and use of negative feedback depending on the age of the students and the context of the interaction. To that purpose, she compared ESL adults' and children's reactions to negative feedback using a task design that involved 32 NS-NNS dyads. The results suggested that both the age of the students and context of the exchanges did have an effect on the interaction patterns. In particular, the adult NS produced more implicit feedback in the form of recasts (a corrective reformulation of a learner's utterance) to their partners than the child NS. In addition, in reaction to feedback, adults appeared to be better able to modify their non-target-like utterances, and adult ESL learners responded to negative feedback more frequently than the ESL child learners in both teacher-fronted lessons and pair-work activities.

For their part, Oliver and Grote (2010) compared several types of recasts provided to child ESL learners and analyzed context (teacher-fronted classrooms versus pair work, child NNS-NNS dyads, and child NNS-NS dyads) and age (comparing their results with those reported by Sheen (2006), who had centered on adults) as well as the impact these two variables had on this feedback technique. They also looked into the learners' uptake (a learner's reaction to feedback which may include a repetition of the feedback, an acknowledgement, a repair, etc.) following these recasts. The authors observed that non-

native children generally provided and received fewer multiple move recasts, but a higher number of single move recasts than adults. Besides, the YLs had a lower level of uptake than adults for all types of recasts in all contexts.

Although still not as abundant as ESL literature, research on child interaction in EFL settings has begun to thrive not long ago, yielding optimistic results which underpin the results obtained in ESL studies to a degree. Actually, children in EFL settings are a fascinating population to study, because FLs (English in its vast majority) are being taught in schools worldwide and this practice seems to be on the increase (Cameron, 2003; Collins & Muñoz, 2016; Enever & Lindgren, 2017; García Mayo & Lázaro-Ibarrola, 2015; Lázaro-Ibarrola & Azpilicueta Martínez, 2015; Murphy, 2014; Pinter, 2007).

An innovator in the study of interaction in an EFL context was Pinter (2007), who carried out a small-scale study in Hungary, where the teaching of FLs is based on mechanical practice such as drilling and memorization rather than on spontaneous and meaningful communication. In particular, she explored two 10-year-old children's ability to interact with each other while completing a spot-the-difference task. She reported cases of peer assistance by providing unknown words, and of children's attention to each other's utterances. In addition, although at the beginning of the study the children showed hesitation and a lack of fluency, as the study progressed, they revealed feeling more confident and were better able to use communication strategies to negotiate meaning.

Among the scant literature on the subject of EFL child interaction, we cannot overlook the investigation carried out by García Mayo and colleagues, framed within a four-year research project funded by the Spanish Ministry of Economy and Competitiveness. For instance, García Mayo and Lázaro-Ibarrola (2015) analyzed the oral production of 80 8-11-year-old children to find differences in amount and type of negotiation of meaning while completing a picture-placement task. The children were enrolled in EFL and content-and-language-integrated-learning (CLIL) programs and were paired to form 40 age- and proficiency-matched dyads (20 EFL, 20 CLIL). The authors reported differences between the two contexts and age groups: On the one hand, CLIL students negotiated for meaning in a higher proportion and used the L1 less frequently than EFL

learners. On the other hand, older learners in both contexts negotiated less and resorted to the L1 more frequently than their younger counterparts.

A more recent study is Lázaro-Ibarrola and Hidalgo (2017a) who also examined the oral interaction of 40 11-year-old CLIL children who worked in dyads to complete a picture placement task. The results revealed that all pairs solved the task successfully using English, although the participants slightly used their L1. In line with previous studies, these YLs used communicative strategies to a lesser extent that adults and ESL children (García Mayo, 2018; García Mayo & Lázaro-Ibarrola, 2015; Lázaro-Ibarrola & Azpilicueta Martínez, 2015). Nevertheless, a more comprehensive analysis of the children's interactions revealed the use of a series of strategies that evidenced their accommodation with each other's communication needs such as completion of each other's utterances, acknowledgment of comprehension of their partner's output and the use of self-repetitions.

The call for longitudinal studies in SLA is an issue which needs to be addressed (Gass & Mackey, 2007). However, it is not always easy to have access to schools, much less to carry out large-scale studies, owing to teachers' tight schedules to cover the government-mandated syllabuses, the time-consuming nature of the research, the difficulty achieving ecological validity or the unlikelihood to secure the presence of the same students in each one of the sessions, to name but a few. The work by García Mayo and Hidalgo (2017) constitutes an example of the changes undergone by the same group of students after one year of instruction. In their study, they analyzed 32 8-10-year-old learners' oral interactions while completing a communicative task twice in two consecutive academic years. More specifically, they examined the use of the L1, the functions it serves, the differences between mainstream EFL and CLIL settings, and also the changes over a year. The results obtained confirm previous research as regards the facilitative role of the L1 when learners have to deal with unknown vocabulary. This use of the L1 to address vocabulary issues was clearly evident in both learning contexts at the two data collection times. In addition, the second time the children carried out the task, the authors observed an increase in the use of the L1 in both learning contexts. Also in line with previous studies, CLIL learners were found to use their L1 significantly less than mainstream learners.

Another recent study within this line of research is García Mayo and Imaz Agirre (2017). They investigated whether age and context (EFL vs. CLIL) constituted factors of influence over the conversational strategies and whether any observed difference would persist over time. To that end, 27 child dyads' oral interactions were examined, and their conversational strategies were analyzed at Time 1 and Time 2, one year apart. Participants were divided into four groups on the basis of their age, 8-9 years old and 10-11 years old, and learning context, and completed two tasks, a picture placement and a guessing game. Their main findings regarding age pointed to a higher use of conversational adjustments and repetitions by the younger learners in both contexts. Concerning educational approach, CLIL learners used fewer conversational adjustments and mainstream children used their L1 to a greater extent. As for the changes over time, the researchers found that conversational adjustments decreased among learners regardless of the learning condition.

Given the paramount relevance of Long's (1996) work, it is no wonder that a large proportion of child SLA research is situated within the interactionist paradigm. Without doubt, progress in our understanding of child interaction in EFL contexts has been made. Generally speaking, we could glimpse that a number of conclusions emerge from the studies discussed so far: (i) An interesting finding in relation to interaction research with YLs is that EFL children also have the ability to interact and negotiate for meaning, although to a lesser extent than ESL children; (ii) these studies seem to be consistent with what has been theoretically put forward in the previous section that both cognitive and social differences exist between children and adults. As a result, the outcome may well be different for children and for adults and so pedagogical intervention must be carefully planned; and (iii) the positive evidence reported on interaction playing a facilitative role on child L2 development.

Definitely, it is an exciting time in child L2 research. Up to now, we have given a rough account of some of the latest research undertaken on one of the most recurring topics concerning child L2 acquisition: oral interaction. However, there remain several aspects of child L2 acquisition about which relatively little is known, and further research is required on these and other questions in order to move the field forward.

One of the constructs which has been under-researched in EFL child SLA, over all in the field of L2 writing, is that of individual differences such as motivation (Azkarai & Kopinska, 2020; García Mayo, 2018; Kopinska & Azkarai, 2020; Kormos, 2012; Waller & Papi, 2017). It is essential to know whether the children's motivation will have an effect on their performance, whether or not their motivation changes throughout the task, and whether or not motivation is task-dependent (García Mayo, 2018). According to Kopinska and Azkarai (2020), given that children's motivation soars and decreases during their school life, more emphasis should be put on gathering learners' beliefs and opinions on the specific tasks they perform in their language classes in order to gain more insights into the motivational processes occurring in the L2 classroom.

As discussed earlier in this chapter, much more research is also needed to fully comprehend the advantages and drawbacks of children learning English at an early age and to understand the similarities and differences of differently aged learners. Empirical studies on educational costs and benefits of an early introduction to language learning are scant, but policymakers impose the teaching of English to millions of children around the world as if benefit was guaranteed. (Copland & Garton, 2014; Oliver et al., 2017).

As compared to research with adults, children's writing is not especially present as a field of inquiry in L2 writing research (Campillo Ferrer, López-Serrano & Roca de Larios, 2012; Manchón, 2011; Ortega, 2009a). Accordingly, García Mayo (2018) suggests another important issue which needs to be dealt with: the *oral-written connection* in child task-supported interaction or, in other words, the impact that child oral interaction has on their written product. As she points out:

> We should not forget that for primary school children, oral and written literacy go hand in hand and the former can facilitate the latter if the teacher uses tasks that enable children to develop their ideas, decide on the language they need to express them and collaborate in organizing them into a coherent written text (García Mayo, 2018, p. 132).

Within this area of inquiry, we would also like to see more research into the potential of collaborative writing (Storch, 2021). Research on SLA has proved that collaborative

writing, which is 'the involvement of two or more writers in the production of a single text' (Storch, 2019, p. 40), is an essential source of learning, because it is through interaction that meaning is created and knowledge is co-constructed (Swain, 2006). Pedagogical collaborative tasks such as information exchange activities, role plays, or pair discussions have become a very helpful tool to engage students in interaction, to provide a context for the negotiation of meaning and, as a result, for all the cognitive processes it triggers (i.e., production of comprehensible input, provision of feedback, modified output, hypothesis testing, etc.). In this manner, by taking advantage of the opportunities these communicative tasks provide, YLs can consolidate their emerging interlanguage (IL). Therefore, if effectively designed and carried into effect, collaborative writing tasks can combine the benefits of speaking and writing and provide rich opportunities for language learning (Storch, 2016). However, we know very little about the effects of children's oral interaction on their joint written product (e.g., Cánovas Guirao, Roca de Larios & Coyle, 2015; Coyle, Cánovas Guirao & Roca de Larios, 2018). For this reason, researchers should also work on designing studies that include a written component (Azkarai, García Mayo & Oliver, 2019; García Mayo, 2018; Storch, 2016, 2021).

Additionally, given that negative feedback constitutes an important part of the interactionist paradigm, future studies would also need to shed light on the debate surrounding the language learning potential of written corrective feedback (WCF). Although a handful of studies has researched the effects of several WCF techniques on writing tasks with adult or adolescent learners (García Mayo & Loidi Labandibar, 2017; Hanaoka, 2006a, 2006b; Hanaoka, 2007; Hanaoka & Izumi, 2012; Kang, 2020; Martínez Esteban & Roca de Larios, 2010; Montealegre Ramón, 2020; Qi & Lapkin, 2001; Yang & Zhang, 2010), research on how primary school-aged children write in collaboration and benefit from WCF is scarce in the EFL context and so would seem to merit further inquiry. Besides, there is a need to compare explicit correction with alternative WCF techniques such as models and reformulations in order to (i) shed light on the inconclusive findings on error correction; (ii) analyze the benefits of those alternative techniques for YLs; and (iii) examine the characteristics of their interaction when dealing with these types of feedback. All things considered, this oral-written connection alongside the provision of feedback can together comprise a highly valuable tool to drive the children's L2 development forward.

Another item which needs to be at the forefront of the research agenda on children is the issue of form versus content in language classrooms. Is it feasible to draw learners' attention to formal aspects of language while teaching content? If so, how could this junction of form and content teaching be accomplished in different contexts? (Oliver et al., 2017). Clearly, one of the challenges in content-based instruction is how to focus on form (FonF) effectively, appropriately and spontaneously in L2 lessons whose overriding focus is on communication (Long, 2000). Recent literature on FonF has looked for ways to call children' attention to linguistic forms, without isolating these forms from their meaningful context. For example, Leeser (2004) suggests that one way to achieve this goal is through the use of collaborative tasks which can encourage children to consciously reflect on their own language use during meaning-based lessons.

Last, but not least, there is also a general call for longitudinal studies on child L2 acquisition that explore the extent to which the findings of these and other issues can be generalized to a wide young population (Bitchener, 2012; Storch, 2018; Manchón, 2011; Oliver et al., 2017). So far, most studies have used small-scale designs, but large-scale empirical research is also clearly needed to observe the development of children's IL, whether variables may change over time, and whether or not our ultimate goal, namely acquisition, takes place.

Certainly, there are many issues of child FL learning that would benefit from further exploration. Given the dearth of studies in this area as compared to the research carried out with adult and adolescent learners, the challenge now is to conduct more research involving YLs, as it is only with sound studies that our results can be robust enough and of help to stakeholders and primary school teachers. In those studies, we should also take into account Pinter's (2014) proposal about doing research with children rather than on children, that is to say, engage them as co-researchers, include them as active participants and pay attention to what they have to say. Although a challenging one, it is also a path full of considerable rewards.

All things considered, the purpose of this dissertation is to answer these calls to the extent possible, thus doing our modest part in, as researchers, enhancing our insights into child L2 acquisition and, as teachers, improving our pedagogical practices as well as our young learners' FL learning opportunities. That is why it is our aim to make all

of these ‘under-construction’ topics mentioned above the core of this thesis project, and therefore they will be treated in more detail in their corresponding chapters.

1.5. Conclusion

In light of the recent trend toward the introduction of English into pre- and primary school levels all over the world, it is becoming extremely difficult to ignore the urgent need for studies on child L2 acquisition, especially in FL contexts. The reason behind this call is that research on child SLA has revealed important social and cognitive differences between children and adults with regard to the processing and learning of an L2. What is more, little do we know about what children really do when they perform specific tasks that facilitate the learning process, and therefore effective pedagogies for teaching YLs in restricted contexts are needed.

Moreover, many reported benefits obtained from studies involving adult learners or high school students have been extrapolated to children and applied as the grounds for teaching methodologies, while studies having to do with children remain scarce. Research on YLs will therefore be advantageous not only to know more about how children acquire an L2 in both ESL and EFL settings, but also to inform policymakers, and to help make decisions about suitable pedagogical practices.

As seen throughout this chapter, research on child SLA has revealed that conversational interaction has a positive impact on L2 learning, so more and more SLA researchers advocate for the use of communicative tasks in second and foreign language contexts (e.g., García Mayo, 2018; García Mayo & Lázaro Ibarrola, 2015; Lázaro Ibarrola & Azpilicueta Martínez, 2015; Lázaro Ibarrola & Hidalgo, 2017a; Oliver, 1998, 2009; Oliver & Mackey, 2003; Oliver et al., 2017; Pica, 2013; Pinter, 2007). It is therefore crucial that the unique characteristics of children along with their individual needs are kept in mind when it comes to designing materials and putting pedagogical approaches into practice (Nicholas & Lightbown, 2008).

Although, from its emergence, research has made a great leap forward as regards child L2 acquisition in FL contexts, there still remain numerous aspects about which

relatively little is known. For example, there is a need to know more about individual differences, or about the extent to which the child's oral interaction has an impact on their written product. Another important line of research is in the area of writing, in particular in collaborative writing, which has been found to be a very helpful tool to engage students in interaction. Related to this issue of writing, we would also like to see more studies on the role that different types of WCF play on L2 learning. To develop a full picture of the potential of conversational interaction, additional studies on FonF in meaning-based lessons would also be needed. Finally, another big challenge for the SLA area is the design of longitudinal studies, since they can provide us with more information about the possible changes that children might undergo over time.

This first chapter has attempted to provide a rough picture of the importance of research on child L2 acquisition and to inform about what is left for us to do. The following pages of this first part of the dissertation will revolve around some of the aforementioned issues which we believe merit further exploration. In particular, the next chapter will focus on the theoretical reasons that support the beneficial effects of WCF on learners' development of linguistic knowledge.

CHAPTER 2: THEORETICAL PERSPECTIVES ON THE LANGUAGE LEARNING POTENTIAL OF WRITTEN CORRECTIVE FEEDBACK

When teachers correct their L2 students' texts, they provide feedback on a number of issues such as content, coherence and cohesion, grammar, spelling, the appropriateness of the vocabulary that is used, etc. However, the type of feedback that has attracted most of the researchers' attention is feedback on linguistic errors. The assessment given to L2 students' production on these specific errors has been commonly referred to as *corrective feedback* (CF) or *error correction* (EC) (R. Ellis, 2009; Van Beuningen, 2010).

Research on feedback is situated at the crossroad of two academic disciplines, both of which have followed different paths in their methodological and theoretical orientations, thus developing almost independently from each other: one set found within the field of L2 writing and another within the domain of SLA (R. Ellis, 2010; Ferris, 2010; Ortega, 2012; Santos, López-Serrano & Manchón, 2010; Van Beuningen, 2010). Researchers in the field of L2 writing have sought to help learners improve the overall quality and efficiency of their written texts and develop their editing and revision skills in a second or foreign language. This perspective has commonly been referred to in the literature as the 'learning-to-write' dimension which perceives feedback as a way of fostering students' competence in producing good quality texts (R. Ellis, 2010; Ferris, 2010; Manchón, 2011). On the other hand, the SLA-oriented domain takes a different stance on CF. Based within a psycholinguistic and cognitive framework, this conception of CF arises from a writing-to-learn strand of theory and research which views CF as highly valued tool to develop leaners' IL (Manchón, 2009, 2011; Ortega, 2009b).

Given that the research focus has recently shifted toward the 'writing-to-learn' dimension and considering that SLA has given much more importance to oral skills than to written skills, the way in which receiving and processing WCF can facilitate learners' linguistic development has not been, until now, particularly visible in the CF research agenda (Santos, López-Serrano & Manchón, 2010). Today, however, much more is known from different perspectives about the value of WCF. Broadly speaking, the language learning potential of L2 writing and WCF can be explained from both social and cognitive dimensions. First, looking at writing through a socio-cultural lens is

grounded on the contemplation of writing as a socially-situated activity, and therefore it considers the innumerable contexts where this activity can take place (Storch, 2016). Second, during the composition of their texts, students engage in a series of mental actions and processes, whose study constitutes the more cognitively-oriented processing dimension of writing research (Manchón, 2018). Given these perspectives of L2 writing viewed from the SLA lens, a wide range of theoretical frameworks has paved the way for research on this area, which helps us distinguish patterns and tendencies in the processes involved in writing in an ESL or EFL context.

Notwithstanding the above, there has been and still is considerable debate about whether WCF is a worthwhile pedagogical practice for L2 acquisition (Bitchener, 2012), since practical and theoretical evidence against its usefulness has been presented (e.g., Truscott, 1996; 2009). Therefore, research has yet to give more conclusive answers as to whether WCF plays a role, where the limits of that role lie, and how it might be most effectively provided to have a significant impact on IL development (Bitchener, 2012). The present dissertation, which falls within the writing-to-learn research strand, aims to contribute to the controversy surrounding the language learning potential of WCF.

Theories, in that they provide explanations for how and why dependent variables (e.g., fluency) are affected by independent variables (e.g., WCF), are the best starting point for a discussion on what the potential might be for WCF to contribute to L2 learning. To this end, in the present chapter the role played by feedback in writing will be clarified by means of the analysis of a number of cognitive and socio-cultural theoretical arguments.

2.1. The role of WCF

If CF has been defined as the assessment given to L2 students' production on specific errors, WCF can be interpreted as the written responses to and comments on linguistic errors in students' written production in a second or foreign language. More specifically, its purpose is to either correct the inaccurate use of the written TL or give

information about where the error has been produced and/or about the reason for the error and how it can be corrected (Li & Vuono, 2019). Recently, WCF has also been operationalized by providing a model text as a means of encouraging learners to identify their own errors or areas of improvement in their written texts (Li & Vuono, 2019).

As mentioned above, the question about whether or not WCF can facilitate language acquisition is still a subject of much debate. Much of this controversy is driven by inconsistent findings and also by the different beliefs on the role that CF plays in both SLA and L2 writing literature (Van Beuningen, 2010). In the field of SLA, the possible effect of negative feedback depends on how the different theories look at it. According to the nativist view, for example, acquisition is motivated by positive evidence, and corrective feedback plays little or no role in acquisition (e.g., Carroll 1995; Krashen 1982, 1985). On the other hand, other researchers have claimed that a mere exposure to positive evidence is not enough for L2 learners to develop native-like accuracy, but provision of negative feedback and attention to form is necessary to drive the L2 forward (e.g., Long 1996; Nassaji 1999; Pica 2002; VanPatten, 1990). For his part, Truscott, who fired the debate surrounding the usefulness of CF due to his fierce opposition to this technique in 1996, suggested that the time and energy spent by both teachers and students on coping with corrections should be more appropriately devoted to other activities, such as additional writing practice. Besides, he questioned teachers' ability to recognize errors, or provide adequate and consistent feedback as well as students' capacity and motivation to use that feedback effectively. It might be the case, however, that teachers' lack of metalinguistic knowledge to provide accurate feedback is likely to be indicative of a need for teacher training rather than a reason for invalidating feedback.

As can be seen in general terms, some researchers believe CF does lead to improvement, and others are more skeptical. What is more, CF opponents have not only affirmed that CF is ineffective to further accuracy development, but some of them have even claimed that it can be counterproductive to the process of L2 acquisition. Some clear examples of this are Truscott (1996) and Krashen (1982) who further contended that CF is not only powerless but also harmful because it causes stress and anxiety of committing the same errors in subsequent written tasks. For these researchers, this anxiety could result in students trying to avoid these errors by producing much more

simplified texts, thus affecting their writing quality. Apart from these practical reasons, Truscott (1996) also brought up theoretical issues that pose a challenge for the benefits of CF. In particular, he alleged that CF overlooks important insights from SLA theory: the role of explicit L2 knowledge in the language learning process and Pienemann's (1989) Teachability Hypothesis. Based on the former construct, CF contestants (e.g., Krashen, 1982, Truscott, 1996) argue that CF can only result in explicit knowledge (if any) which, according to Krashen (1982), is rather limited and can only be used during monitoring (i.e., editing of output), while online language use is completely motivated by implicit knowledge. Given Krashen's idea that explicit knowledge can never become implicit, Truscott (1996) came to the conclusion that CF can only lead to 'a superficial and possibly transient form of knowledge' (p. 345).

The latter theoretical argument raised by Truscott (1996) is based on the Natural Order Hypothesis (Krashen, 1981) and Pienemann's (1989) Teachability Hypothesis which argue that no intake (input which is comprehended and that impacts the learner's developing linguistic system) can take place if students are provided with a list of corrections that they are not yet ready to acquire, because the acquisition of an L2 is supposed to follow a fixed sequence that is resilient to external factors such as feedback. In Truscott's (1996) view, for CF to have any potential impact, it should be aligned to the learner's current level of L2 development. It has been highlighted, however, that what is known so far on developmental sequences is still too scant to be helpful for teaching practice (e.g., DeKeyser, 1998; R. Ellis, 1997). This led Truscott (1996) to conclude that not even provision of adjusted CF is beneficial for the moment and cannot therefore be expected to be of help to SLA.

Truscott's absolute rejection of (W)CF has received much criticism from scholars in writing research (e.g., Bitchener, 2008; Ferris, 2004; Sheen, 2007). From a theoretical standpoint, SLA considers that the teaching of a language should be meaning-focused. On the other hand, a focus on formal aspects by means of explicit instruction and CF is nowadays becoming widely accepted as a prerequisite for the learning of non-salient features (such as the English third person singular morpheme -s) that can go unnoticed especially by our younger learners (Van Beuningen, 2010). Empirically, as will be discussed in later chapters, research has unmistakably proved that WCF does foster L2 development as evidenced in learners' subsequent pieces of writing (Kang & Han,

2015, Li & Vuono, 2019). On the pedagogical front, applying feedback on learners' errors is typical of writing classes, and studies have demonstrated that teachers, students, and other stakeholders are all positive about WCF (Chen, Nassaji, & Liu, 2016; I. Lee, 2009; Li & Vuono, 2019).

The theoretical rationale supporting the view that WCF can be of help for L2 development rests on both cognitive and sociocultural constructs on learning. Cognitive theories integrate leading SLA concepts and hypotheses such as the focus-on-form paradigm (Long, 2000), Skill Learning Theory (DeKeyser, 1997), the Output hypothesis (Swain, 1985, 1995, 1998, 2000, 2005) or the Noticing hypothesis (Schmidt, 1990, 1994, 2001) among others. Sociocultural theory (Vygotsky, 1978) has also furthered our insights into the language learning potential of WCF by means of the significance adhered to cognitive development as a socially situated activity. This idea is rooted on the Vygotskian premise that the acquisition of an L2 can be co-constructed through collaborative dialogue during problem-solving activities. In what follows, these theoretical constructs will be considered in greater detail in an attempt to examine the key issues surrounding the beneficial effect of WCF on students' linguistic development.

2.2. Cognitive theories and WCF

Cognitive theories in the field of SLA have attempted to explain the processes involved in the acquisition of an L2. For many years, the research lens has been placed on oral communication, considering writing as an end product of acquisition, rather than as a means to an end (Bitchener, 2012). As of late, however, more and more SLA researchers have put forth a solid defense for the affordances of written communication, particularly pertaining to WCF (Bitchener, 2012; Manchón, 2011; Polio, 2012). The belief that writing in a second or foreign language is too intricate and cognitively demanding to be introduced into the primary school language classroom has therefore been ruled out and it is now increasingly accepted that, as with adults, helping children to express themselves through writing might potentially foster language development (Manchón, 2011). It is through writing and the provision of feedback on their final

drafts that children activate such cognitive processes as hypothesis formation and testing, noticing, metalinguistic reflection and problem-solving strategies, all of which may lead them to reassess their existing L2 knowledge (García Hernández, Roca de Larios & Coyle, 2017).

The reasons underpinning the facilitative role of writing are based on the idea that learning occurs precisely due to the essential qualities of the writing process (Bitchener, 2012; Polio, 2012). In this sense, the off-line nature and the visual permanence of both the written product and the feedback provided ensure a more in-depth linguistic analysis, as students would have more time to focus on language both during and after the written task (Cánovas Guirao, 2017; Kormos, 2012). The absence of time constraints can also promote the noticing of mismatches between the learners' own written output and the feedback received, as well as the production of more accurate form-function mappings (Cánovas Guirao, 2017). Therefore, the time- and visual-related aspects inherent to written communication would seem to be paramount to the language learning potential of WCF, since they result in visible, (relatively) permanent output, more time for content planification, linguistic encoding and revision processes, all of which could potentially result in the use of more complex language, as compared to what would be feasible in spoken language (Vasylets, Gilabert & Manchón, 2019).

2.2.1 The interaction model

Of all the cognitive approaches that have something to say about the impact of WCF in L2 acquisition, the Interaction Hypothesis (Long, 1996) is the most complete and has possibly the most to offer (Bitchener, 2012). According to Long (1996), learners' engagement in conversational interaction facilitates SLA. The constructs born from the interaction model which are believed to be helpful for L2 learners include being exposed to language (input), producing language (output), negotiating meaning and form, and receiving feedback on that output which, in turn, pushes them to modify their output during conversation. These processes provide learners with opportunities to notice differences and similarities between their output and the target-like forms supplied in the input and to determine where discrepancies lie. In Long's (1996) view, while interaction cannot explain the whole process of L2 learning, it is a necessary condition for learners to acquire L2 communicative competence. This hypothesis

subsumes some aspects from Krashen's (1982, 1985) Input Hypothesis, Swain's (1985, 1995) Output Hypothesis, and Schmidt's (1990, 1994, 2001) Noticing Hypothesis which concentrate on the importance of input, output and noticing (respectively) in language acquisition. Although research on the mechanisms that mediate between them is still on-going, it is now widely recognized that there is a solid relationship between interaction and learning (Adams, 2007; García Mayo & Alcón Soler, 2013; Loewen & Sato, 2018; Long, 1996; Mackey, 2007; Mackey & Goo, 2007).

Having outlined the basic tenets relevant for the interaction approach to SLA, we turn now to a review of each of the key constructs necessary for understanding how interaction and learning are related.

2.2.1.1 Input

Input refers to the language learners are exposed to through any medium (listening, reading or gestural in the case of sign language) (Gass & Mackey, 2007). All theories of SLA acknowledge the importance of input as a key component in the acquisition process (Gass & Mackey, 2007; Loewen & Sato, 2018), but they differ as to how much input is needed for this process to occur and how it should be organized (Gass & Mackey, 2006). Krashen's (1982 *et passim*) Input Hypothesis states that learners progress in their knowledge of the language when they comprehend language input that is slightly beyond their current level of competence. Krashen called this level of input 'i+1', where 'i' is the learner's IL and '+1' is the next immediate step along the development continuum of language acquisition. Put it differently, he affirms that exposure to comprehensible input alone is sufficient for L2 acquisition. However, his proposal has not been without criticism since, among other issues, some of his constructs have been considered vague, imprecise and not easily testable. More importantly, the Input Hypothesis concentrates exclusively on comprehensible input as sufficient for acquisition to the neglect of any possible intervention of output (Gregg, 1984; McLaughlin, 1987).

2.2.1.2 Output as a noticing facilitator

As opposed to input, *output* is the term used to refer to the language that learners produce in the oral and written modes (Gass & Mackey, 2007). For Krashen (1982), output was merely a product of already acquired L2 knowledge. However, years later, Swain (1985, 1995, 1998, 2000, 2005) proposed the Output Hypothesis, which sought to rectify the inadequacies of the Input Hypothesis by maintaining that language acquisition is also possible through the production of language. Swain also argued that the significance of L2 output lies in the fact that it pushes learners to process language more deeply (i.e., beyond semantic processing) and with more mental effort than is necessary during exposure to language. Thus, in her view, output constitutes not only the result of L2 learning, but also a fundamental step in the process. In her Output Hypothesis, Swain (1993) identified several functions of output in the process of L2 learning. These include engaging students in opportunities to (i) formulate and test hypotheses about how the TL works and receive subsequent feedback on these hypotheses; (ii) generate metalinguistic reflection that enables learners to understand the relationship between meaning, form and function; (iii) develop fluency and automaticity, since output facilitates production and practice; and (iv) notice the gap between what they want to say and what they are able to say, which may result in learners' conscious recognition of their linguistic problems. The Output Yypothesis, therefore, complements and addresses the deficiencies of the Input Hypothesis by emphasizing the importance of meaningful language use both in speaking and writing activities.

According to Cánovas Guirao et al. (2015), from these arguments it would seem plausible to think of WCF as a prompt for the successful allocation of learners' attentional resources during output activities. This view was supported by Swain (1991), who admitted that output by itself would not be sufficient to serve the aforementioned functions. In her own words, 'if students are given insufficient feedback or no feedback regarding the extent to which their messages have successfully (accurately, appropriately, and coherently) been conveyed, output may not serve these roles' (p. 98). Other researchers have also supported this idea that in order to be favorable for the language learning process, learners' output should go hand in hand with CF:

> While the focus is on meaning, there is a limit to how much an L2 learner can introspect the sufficiency of his own linguistic resources. Also, even if the learner consciously recognizes at that point what he lacks, there is no guarantee, for various reasons, that he will subsequently be able to tune himself in for a solution in the future input, or even if he is, he may not be able to tell whether what he sees as the potential solution is actually the correct solution. Rather, external feedback [...], I shall argue, may significantly facilitate the fulfillment of the 'noticing' function (Han, 2002, p. 18).

Written output, then, is believed to trigger the noticing of linguistic aspects that are not present in students' IL (holes), alongside those they have partially acquired (gaps) generating a need to attend to relevant input to solve them (Izumi, 2013). The limitations and difficulties that students undergo during written production may cause them to re-evaluate their knowledge of the L2 through processes of hypothesis testing and metalinguistic reflection (Swain, 1985 *et passim*). Accordingly, WCF has the power to promote the noticing and consolidation of new or existing knowledge as well as to enhance consciousness of gaps between their output and the TL (Han, 2002). In this sense, writing may foster L2 development by generating a loop between output and input in the form of feedback via processes of noticing (Yang & Zhang, 2010). In addition, as mentioned in the previous pages, the positive influence of written output and WCF would seem to overshadow that of oral production and correction, since the visual permanence and the lack of time constraints as opposed to the fleeting nature of spoken language would allow for a better exploitation of the feedback.

2.2.1.3 Noticing and WCF

Over the last two decades, SLA research has placed much emphasis on the role of attention, awareness and noticing in triggering learners' IL change (e.g., Robinson, 1995; Schmidt, 1990, 1995, 2001; Schmidt & Frota, 1986; Swain, 1985, 1995). The importance of these concepts was such that we can find in the early literature statements such as ''those who notice most, learn most'' (Schmidt & Frota, 1986, p. 313) and ''no

noticing, no acquisition'' (R. Ellis, 1995, p. 89). Later, Hanaoka (2006b) suggested that 'noticing is a prerequisite for L2 learning to take place' (p. 167).

Traditionally, research on noticing has been associated with input. Consequently, noticing has been referred to as the intake of features that occurs when learners pay conscious attention to input and described as the first stage of language acquisition (Batstone, 1996). Nevertheless, as Swain (1985) points out, there are forms of intake that do not derive from input but are generated in output and that may also be facilitative of L2 acquisition. Later on, Schmidt (1990, 1994, 2001) proposed the Noticing Hypothesis, which claims that L2 learning does not happen without awareness, and that it is only through conscious attention to specific features of the L2 during input and output processing that language learning can take place. Schmidt (1990) distinguishes two levels of awareness: awareness at the low level (noticing) from awareness at the high level (metalinguistic awareness or understanding). While noticing is the perception of surface level features in output and input/WCF, understanding is associated with learning at a deeper level. Although a higher level of awareness may not be a prerequisite for SLA, Schmidt maintains that learning is not likely to occur without some degree of consciousness. However, the Noticing Hypothesis has found disagreement on two main grounds. First, the constructs of noticing and understanding are not clearly defined and have been questioned by defenders of the idea that awareness plays no role in L2 acquisition or, in other words, that language acquisition does not require awareness (Schmidt & Frota, 1986; Tomlin & Villa, 1994; Truscott, 1998). Second, other researchers have also proposed that a certain level of understanding is necessarily involved in the registration of linguistic exemplars in the learner's developing system (Truscott, 1998; Truscott & Sharwood-Smith, 2011). Notwithstanding the above, the weaker version of Schmidt's (2001) hypothesis (i.e., noticing does not necessarily result in language learning, but it can certainly facilitate it) has attracted considerable support from numerous studies and is nowadays widely accepted (Izumi, 2013).

Noticing has also been used to explain the problem-solving processes that go on during written production, when students search for the linguistic means necessary to convey their intended meaning (Manchón & Roca de Larios, 2007). The involvement of noticing in the analysis of the input and output processes undergone in SLA has led to a

distinction of two different types of noticing which are helpful to understand how learners manage to convert input into intake during written production and input processing. According to Izumi (2013), one type of noticing occurs during output production and refers to the idea of noticing a gap between what students want to express and the internal realization that they do not have the means to convey the message fully or appropriately. While Swain (1998) uses the terms *gap* and *hole* interchangeably, Izumi (2013) identifies a subtle difference between the two of them. In his view, both refer to internal processes, but the latter entails a complete absence of the linguistic feature in the student's repertoire, while the former suggests that the student is able to express the concept partially and imprecisely, which could result in an internal search for a solution when trying to express the message. Applied to the process of L2 writing, 'noticing the hole' would show up as *covert* problems (those difficulties in output production which are not addressed), and so we would expect the use of compensatory writing strategies such as message abandonment or message replacement due to the student's lack of linguistic resources (Roca de Larios, 1999). On the contrary, 'noticing the gap' is generally associated with the existence of *overt* problems (difficulties that leave observable traces in learner's spoken or written output (Hanaoka & Izumi, 2012), that are usually addressed through strategies such as word approximation (an alternative lexical item that shares the target word's semantic features) or signaling awareness (an indication to the reader that the chosen solution is unsatisfactory) (Cánovas Guirao, 2017; García Hernández, 2017).

According to Schmidt (1990), the 'gap' must be consciously addressed so that input can be turned into intake. During input processing, however, partial learning can occur if learners notice only part of a linguistic form, (for instance noticing a noun without its collocations). For Hanaoka (2007), learners' prior L2 knowledge is what affects this phenomenon, which she denominates the 'scope of noticing', in such a way that if the feature present in the input was completely new to the learner, they would be less likely to notice the gap, and they would also need to start a new noticing of the relationship between form, function and meaning. On the contrary, if the feature was partially known, then the learner would only have to learn the additional linguistic information, and the possibilities of noticing the gap would be higher. R. Ellis (1995), however, prefers the term 'cognitive comparison' rather than noticing the gap since, to him, learners also need to notice when their output is similar to the input they receive. For

instance, as he sees it, noticing similarities and differences between their own output and WCF can bring about destabilization and restructuring of existing L2 knowledge through hypothesis testing (Gass, 1997; Long, 1996).

What is most important is that by noticing that there is something they cannot express or can express incompletely, learners are made aware of their production problems and thus pay heed to subsequent input in a selective way which in the written mode would mean directing their attention to WCF (Manchón, 2011). Research on WCF has actually confirmed that learners notice holes during their first compositions, autonomously search for solutions in the input provided in the form of feedback, and incorporate them into subsequent revisions (e.g., Cánovas Guirao et al., 2015; Hanaoka, 2007; Luquin & García Mayo, 2020; Yang & Zhang, 2010). Furthermore, when learners notice the differences and similarities between their IL forms and the target-like forms, they are believed to engage in processes such as matching (Klein, 1986) or the already mentioned cognitive comparison (R. Ellis, 1995). These processes, along with hypothesis testing, are deemed crucial for the continuous assessment of the learner's internal systems and, eventually, for the acquisition of the new knowledge (Gass, 1997).

Furthermore, we should also point to the advantage of WCF over orally provided feedback when it comes to noticing, because, although both modes give learners the opportunity to notice a mismatch between the TL and the IL, students are less likely to make the cognitive comparison in online oral interaction (Abe, 2008; R. Ellis, 1995; Hanaoka, 2006a). In addition, when presented with a considerable number of stimuli, the human brain might not be able to allocate their attention to all of them due to a lack of available processing capacity (Schmidt, 2001). The challenge of dealing with spoken language, oral CF and attentional resources altogether may cause a cognitive burden. During the writing stage, on the contrary, the availability of time allows students to engage in a cognitive comparison between their output and the feedback provided for an extended period of time, thus increasing the probability of their noticing gaps in their IL (Polio, Fleck & Leder, 1998). Therefore, a greater amount of noticing is expected in writing.

Apart from the amount of time available, Hanaoka (2006b) observes two more related differences between written and oral feedback. The first one has to do with the

provision of target-like forms which are referred to as *recasting* in oral interactions and *reformulation* (the rewriting of the whole text produced by the learner) in writing. While the former 'only' corrects local errors in an utterance, the latter can address difficulties 'at the textual as well as sentence levels' (Qi & Lapkin, 2001, p. 282). Relatedly, the second difference refers to the learner perception of the purpose of the feedback. Recasts, on the one hand, serve multiple functions, and their corrective goal may not be noticed by learners. As a matter of fact, numerous studies have shown that this feedback technique is sometimes too ambiguous to be noticed (e.g., Lyster, 1998; Lyster & Mori, 2006; Lyster & Ranta, 1997). On the other hand, reformulations push students to find differences between their initial output and the received feedback. Consequently, the purpose of the feedback might be clearer to them. Given the differences stated above (time factor, sentence- versus discourse-level feedback, and learner perception of the purpose of the feedback), Hanaoka (2006b) concludes that noticing that takes place in response to WCF deserves more attention in its own right.

Certainly, there is a lack of empirical research studies on the role of noticing in L2 writing. Questions such as how noticing is connected to L2 composing and what effect it has on L2 writing improvement still need to be addressed and studied in more detail (Hanaoka, 2006a; Qi & Lapkin, 2001), even more so with primary school-aged children.

2.2.1.4 Feedback in writing

Although we have been talking about it indirectly throughout this chapter, equally important for the interactionist perspective is the feature of feedback, which, as indicated earlier, has been an issue of debate in both L2 writing research and pedagogy. Those scholars who argue against feedback on learners' writing maintain that it is inefficient, and the revisions may de-motivate learners, since they could see themselves as weak writers. In addition, WCF contestants add that no matter how much feedback students receive on their writings that they will keep on committing language mistakes in their subsequent compositions, so WCF does not even guarantee long-term effects on learners' accuracy in writing (Polio, Fleck, & Leder, 1998; Truscott, 1996).

On the contrary, those in favor of feedback in writing (e.g., Bitchener, 2008; Ferris, 2010) point out that although feedback may not prevent students from committing errors, it can raise their awareness of correct forms. In this sense, it seems clear that, even with a high proficiency level, sudents cannot avoid language errors when writing. Therefore, we cannot expect them to notice gaps or experience difficulties without the mediation of teachers' or peers' feedback.

Just as input constitutes the positive evidence that learners need to construct their second language grammars, feedback is the negative evidence necessary to amend those errors that L2 writers inevitably commit with the cycle of interaction repeating until a hypothesis is confirmed (Loewen & Sato, 2018). When learners' wrong hypotheses and ill-considered generalizations result in errors in their written output, feedback from the teacher is needed in order to help students correct these errors (Qi & Lapkin, 2001). Therefore, appropriate feedback helps learners gain consciousness of their holes/gaps in writing, leading at the same time to progress in their future written output based on the feedback received. Contrary to what WCF opponents contend, significant and long-term improvement in writing requires learners to start with particular texts and to go through short-term revisions (Ferris, 2002).

In sum, the role of WCF as a facilitator of L2 learning cannot be explained without the help of the interaction approach, which brings together the cognitive processes that take place in conversational interaction. The basic tenet of the interactionist perspective is that input and interaction provide language learners with opportunities to notice differences between their IL and the TG. They also receive feedback on their production which in turn pushes them to modify their own output during conversation.

Interaction plays a significant role in SLA theory and pedagogy, and there seems to be no slowdown in this respect. Accordingly, research on interaction will need to further examine how to best implement interaction in the language classroom so that L2 learners can obtain the greatest benefit for IL development and communication skills. As was mentioned in the previous chapter, a large amount of scholarly work on interaction to date has focused on adult or adolescent ESL populations, and results from these contexts cannot be extrapolated to ESL children nor to EFL settings to determine

pedagogical practices (García Mayo, 2018). In the first chapter, we reviewed some examples of research on interaction conducted with EFL children but, certainly, more studies are needed to understand the relationship between conversational interaction and learning, and even more so between interaction and L2 writing.

The following section cannot be considered a cognitive theory in itself, but it is certainly framed within it. It draws from one of the constructs of the Interaction Hypothesis, namely, attention, and explains part of the potential that WCF has on language learning.

2.2.2 Writing as a Focus on Form intervention

As clarified in the previous section, a great amount of SLA research supports the idea that effective L2 pedagogy should involve attention to linguistic form. In its absence, L2 acquisition is assumed to be slower, more difficult, and less successful (Doughty, 2003). *FonF* is a central construct that emanates from the Interaction Hypothesis. Actually, the term was first introduced by Long (1988, 1991) himself to refer to an approach or a pedagogical intervention that directs students' attention toward linguistic elements in content-based lessons and whose purpose is to prompt noticing. In Long's (2000) own words, FonF 'involves briefly drawing students' attention to linguistic elements [...] in context as they arise incidentally in lessons whose overriding focus is on meaning or communication. The temporary shifts in focal attention are triggered by students' problems with comprehension or production' (p. 185).

The concept was initially conceived of as an alternative to the *focus on forms* (FonFs) paradigm, where particular linguistic features are taught directly and explicitly (Long, 2000). The scope of the term 'FonF', however, has stretched over time, adopting a broader perspective (R. Ellis, 2016). Among some of the refinements of the definition we find R. Ellis' (2015), who claims that the term is best understood not as an approach, but as a set of activities or procedures. Understanding FonF as a series of activities leads to another expanded characteristic: although Long's definition implies that FonF episodes are unplanned, for some researchers they can be both planned and unplanned, and reactive as well as pre-emptive (R. Ellis, 2016). In addition, it should be pointed out that although it usually refers exclusively to grammar, form can actually refer to any

aspect of language: lexical (phonological and orthographic), grammatical or pragmalinguistic (R. Ellis, 2016). However, in the present dissertation we will mainly address form as grammar.

2.2.2.1 The need for FonF

Current interest in FonF is somewhat influenced by the findings observed in immersion and naturalistic settings, where the only foci were on meaning and communication. Research conducted in French immersion programs in Canada (Harley & Swain, 1984; Swain, 1985) revealed that students had developed strategies to make themselves understood, but the messages lacked grammatical accuracy; that is, the formal side of language was left unattended. In addition, learners did not receive enough feedback in this type of contexts, and opportunities for output were very scarce (Pica, 1994, 2002).

Another reason that brought about the rejection of a pure focus on meaning in L2 learning is the significance of noticing. Some attention to form is essential for language learning. '[...] leaving learners to discover form-function relationships and the intricacies of a new linguistic system on their own makes little sense' (Doughty & Williams, 1998, p. 11). In other words, we cannot expect learners to move from no knowledge of the linguistic rule whatsoever to mastery without formal instruction.

Moreover, Pienemann's (1989) claim that students will only be able to comprehend and produce those L2 forms for which they show developmental readiness has also motivated the return to some kind of grammar instruction. If learners try to use a structure before it has been acquired, they may commit predictable errors (Pienemann, 2007). In both naturalistic and classroom settings, students hardly ever manifest immediate acquisition of new forms, but it is conditioned by fixed developmental sequences (Long & Robinson, 1998).

On the contrary, from a FonFs intervention, it was observed that the downside of decontextualized grammar instruction is that students have problems transferring the knowledge they have gained from isolated grammar lessons to actual language use in a communicative task (Van Beuningen, 2010). Although the FonF paradigm was

designed to ‘patch’ the problems found in both focus on meaning and FonFs contexts, one of the challenges that has concerned researchers and teachers is how to focus on form in a way that is both effective and appropriate without isolating these forms from their meaningful context (Leeser, 2004).

Let us now consider some of the solutions that researchers have found and investigated over the years to carry out FonF-oriented language lessons successfully.

2.2.2.2 Types of FonF

As we have explained, a key concern in instructed SLA research and pedagogy is how learner attention can be directed to linguistic forms during L2 learning communicative tasks. The quantity and the quality of input in FL contexts is not enough/adequate for learners to pay attention to relevant target forms and that is why some type of intervention (focus on form) is necessary. Although this issue has been the subject of many investigations in the course of the last two decades, researchers have only more recently set about looking into ways in which learner’s attention can be drawn to L2 features (Lee & Révész, 2018). Considering that attention serves a fundamental function in mediating the process of selecting input for further processing (Robinson, 2003; Schmidt, 2001), scholars have exhibited an enthusiastic concern about discovering the extent to which several teaching techniques can attract attention to L2 constructions. One way to draw students’ attention to linguistic forms is through input enhancement (R. Ellis, 2016).

(a) Input enhancement

Research on input enhancement has shown that perceptual saliency facilitates learners’ noticing of the target-like forms (Doughty & Williams, 1998; Lee & Révész, 2018). Salience of the input allows students to notice formal aspects in the input, in such a way that the more salient the input, the more opportunities to advance students’ L2 knowledge (Meguro, 2019; Sharwood-Smith, 1991). In writing, the saliency of the target forms is usually marked by using italics, bold, capital letters, increased font size, or change of font.

Nevertheless, R. Ellis (2016) cautions that noticing is dependent on a number of factors. One of these factors is the nature of the target form itself, because some structures are more salient than others (for example, third person singular morpheme '-s' is not as perceptible as other features such as articles or lexical forms). It is also more likely that students allocate their attention to an enhanced form if they have in part acquired it and/or have some knowledge of it than if it is completely new to them, an issue which reminds us of the aforementioned Hanaoka's (2007) 'scope of noticing'. However, the factor that most seems to hinder learners' perception in R. Ellis' (2016) view is the learner's proficiency level. In general terms, low-proficiency students find it more difficult to simultaneously grasp the meaning of the text and consciously attend to linguistic form and are more likely to prioritize meaning over form. In this regard, some research (e.g., Izumi, 2002; Shook, 1999) has demonstrated that input enhancement has no effect on reading comprehension. Unless it is combined with other teaching techniques that prompt intentional learning, students are likely to engage in top-down processing and pay little or no attention to the highlighted items when completing a comprehension task. S. Lee (2007), however, reported reversed results, that is, whereas input enhancement helped students learn the passive forms, participants in the enhanced conditions remembered significantly fewer idea units than those in the unenhanced conditions. In a similar vein, although Wong's (2003) study could not prove its benefits on L2 learning, text enhancement enabled students to remember more easily the highlighted information.

Lee and Jung (2021) investigated this issue in greater depth. They analyzed the extent to which textual enhancement, task manipulation and their interaction would have an impact on 73 Korean college students' attentional processing and development of English participle phrases that were typographically enhanced using a different color. To this end, the participants read an article in one of four experimental conditions: (a) textually enhanced, careful reading, (b) textually enhanced, expeditious reading, (c) textually non-enhanced, careful reading, and (d) textually non-enhanced, expeditious reading. While performing the task, the learners' eye-movements were recorded with an eye-tracker to measure the amount of attention paid to the target construction. The findings indicated that both textual enhancement and task manipulation impacted positively on the learners' attention to the target construction during reading, but

perceptual saliency did not lead to L2 grammatical knowledge development. In view of these results, the authors assume that the learners opted for directing their attention to meaning over form, as they had to make a great cognitive effort to comprehend the content of the text.

A recent longitudinal study by Chung and Révész (2021) investigated the impact of textual enhancement in post-reading tasks on the use of the third person singular *-s* morpheme by child learners. The 49 learners participated in task-based reading lessons in their own classrooms, and they were randomly divided into two groups, one being exposed to textual enhancement and the other not. Pretest-posttest development was assessed with a grammaticality judgement test. The findings showed a positive effect for textual enhancement, which the authors ascribed to (i) the use of a longitudinal design, (ii) the incorporation of textual enhancement into the post-task rather than the during-task stage, (iii) age of participants, and (iv) prior knowledge. What is more, the children showed a very good understanding of the readings irrespective of whether or not they received textual enhancement. In consequence, it seems that there are no clear answers yet as to how enhancing techniques would influence L2 reading processing and learning. It is, thus, necessary to continue exploring these issues when learners receive input enhancement through writing.

(b) Corrective feedback

Another pedagogical tool identified as a potential focus-on-form instrument is indeed CF (e.g., R. Ellis, 2005). As we have been explaining throughout the present chapter, CF constitutes a reactive FonF methodology that takes place in a context of negotiation of meaning and form (R. Ellis, 2016), and it is used to induce students' attention to form while completing a task in a personalized and individualized way (Van Beuningen, 2010). As pointed out by R. Ellis (2016), no type of FonF has received more attention than CF, but this is true only of oral CF. As previously stated, research on written CF is not as abundant and its interest lies partly in the fact that CF on written output is argued to be particularly promising as a focus-on-form intervention due to its off-line and permanent nature. Whilst oral CF will inexorably interrupt the flow of communication, students only have to cope with WCF after meaning has been communicated (Polio, Fleck, & Leder, 1998).

It is widely accepted that writing alone can constitute the perfect context for the study of non-directed forms of noticing and FonF processes and, as such, it is an ideal scenario for the investigation of self-initiated FonF processes (Santos, López-Serrano & Manchón, 2010). Thereby, it represents a possible site for learning since, as pointed out by Williams (2001), 'the effectiveness of FonF is ultimately determined by learner need' (p. 175). This would account for the analysis of those language use contexts (formally termed *language-related episodes*, see section 2.2.2.3) in which students decide to focus on formal aspects (Williams, 2001). In a complementary way, feedback processing potentially facilitates learner's engagement in actions (such as noticing and FonF) which may derive in learning effects. As clearly worded by Sheen (2010, p. 175):

> instead of viewing the goal of teaching writing as that of improving the learners' writing skills, practice in writing can be seen as one form of output that in conjunction with CF can facilitate interlanguage development. In other words, instruction that incorporates written CF constitutes a technique to draw L2 learners' attention to linguistic forms in their own output and thereby facilitate acquisition.

(c) Pre-task planning

A further factor that may guide learners toward a FonF process is pre-task planning (R. Ellis, 2016). Research on pre-task planning makes a distinction between 'guided planning' where students' attention is attracted toward a specific aspect of language or even a particular grammatical item, and 'unguided planning' where they are left to decide for themselves what aspects they plan (R. Ellis, 2016). In writing, pre-task planning would be translated as instruction on how to best exploit the feedback provided (Cánovas Guirao, 2017).

Students may be well used to receiving and analyzing feedback in ESL and EFL contexts where FonFs and writing tasks are commonplace. However, this is generally the case with older learners who are usually more willing to pay attention to information given on their linguistic errors. Written tasks and feedback provision on younger learners' written output tend to be, on the contrary, disregarded in favor of oral tasks.

This issue justifies why children today are often unacquainted with writing in a second or foreign language, but also why they are so disoriented when trying to decipher the feedback received on their texts. On that account, pre-task planning, or alternatively, pre-task preparation, designed and intended for aiding YLs in identifying and comprehending the nature of their errors could conceivably strengthen the quality of their noticing and processing of WCF (Cánovas Guirao, 2017; Yang & Zhang, 2010).

A number of studies have actually pointed to the need for students to be instructed to notice mismatches between their written output and the feedback provided in an attempt to help children handle feedback properly (Allwright, Woodley & Allwright, 1988; Barnawi, 2010; Cánovas Guirao, 2017; Cánovas Guirao et al., 2015; Coyle et al., 2018; Luquin & García Mayo, 2020; Qi & Lapkin, 2001; Yang & Zhang, 2010). For example, Allwright et al. (1988) proposed that letting students jointly examine their compositions and the feedback received through class discussion may be even more beneficial than the particular type of feedback itself in showing L2 writers how to modify their texts correctly and raise awareness of their language holes. For their part, Qi and Lapkin (2001) also agreed on the fact that teachers should train learners, especially those with a low proficiency level, to notice gaps through awareness-raising activities. Yang and Zhang (2010) also joined this initiative and stressed the importance of the teacher's intervention during feedback comparison tasks in the form of whole-class discussions of the feedback provided to help focus students' attention on particular linguistic aspects. Cánovas Guirao et al. (2015) suggested that younger learners need 'extended practice at the feedback comparison stage in activities designed to promote noticing and rehearsal so as to facilitate children's encoding of linguistic forms in long-term memory for future retrieval and use' (p. 73). Studies on pre-task planning and its effects on noticing processes, L2 writing development and learning over prolonged periods of time are almost non-existent. A notable exception is the recent work done in an EFL context by Cánovas Guirao (2017), who devoted six weeks to training the treatment group (TG) to work with models through guided class discussion and actually reported positive results for those who had benefitted from instruction against those who had not.

Everything considered, more studies are clearly needed to understand how concrete planning strategies influence performance of a task, noticing processes and development of the writing skill in the short and in the long term.

(d) Task repetition

According to R. Ellis (2016), task repetition functions in much the same way as pre-task planning in that the first performance (or the pre-task session) allows learners to reflect on what to say and how to say it. Afterwards, it allows them to pay attention to form in the repeat performance. Task repetition is a technique which is supposed to raise students' attention to both meaning and form in efficient, effective and accurate ways (Ahmadian, 2012). The main claim in the SLA literature is that when students are asked to perform a task for the first time, their main focus is on getting familiar with the task and on conveying meaning, whereas in the second performance, where the cognitive burden is lower, learners can focus their attention more closely on formal aspects of the language (Bygate, 2001, 2018; García Mayo, 2018; Hidalgo & García Mayo, 2019). This means that language use activates a number of processes and when learners are provided with the opportunity to cope with the same or a similar task, different language dimensions will be attended to and, as a result, different aspects of the student's first performance are likely to change (Bygate, 2001, 2018). Therefore, task repetition represents a favorable context for FonF.

Although there are several studies that corroborate the beneficial effects of task repetition on L2 adult learners, few studies to date have worked with child populations even rendering somewhat conflicting results, probably due to the great diversity of the variables analyzed (context, age, level, tasks, and time intervals) (Azkarai et al., 2019; R. Ellis, 2016; Lázaro-Ibarrola & Hidalgo, 2017b; Sample & Michel, 2014). For instance, it seems that differences exist regarding task repetition between ESL and EFL settings during oral interaction. Azkarai and Oliver (2019) observed that task repetition helped young EFL learners commit fewer errors, whereas young ESL learners delivered more recasts and negotiated for meaning to a higher extent. Therefore, although task repetition in both ESL and EFL settings would seem to be positive, its benefits may still vary according to context.

In relation to writing, a recent study by Hidalgo and Lázaro-Ibarrola (2020) focused on the effects of task repetition in conjunction with collaborative writing on the performance of 10 12-year-old EFL pairs. To that end, the authors analyzed the pair talk

and the three texts produced by the students in response to the same visual prompt by using quantitative (complexity, accuracy and fluency (CAF)) and holistic measures. Results revealed that upon task repetition, the children's texts improved when measured holistically, most language-related episodes were form-focused, and their number decreased with each repetition. In view of these findings, and contrary to studies showing that task repetition has an impact on the CAF of learners' oral performance (García Mayo, 2018), the authors support the use of holistic measures when analyzing learners' written production and also point to the benefits of collaborative writing in the context of task repetition with children.

From a task-type perspective, the results reported in the literature suggest that those tasks in which both content and procedure (*exact task repetition*) are repeated result in greater accuracy, while fluency seems to be favored by tasks in which only the procedure (*procedural task repetition*) is performed again (García Mayo & Imaz Agirre, 2016; García Mayo, Imaz Agirre, & Azkarai, 2017; Lázaro-Ibarrola & Hidalgo, 2017b). Hu (2018), for instance, also used CAF measures to analyze the participants' production and found that exact task repetition facilitated the accuracy of the EFL adult learners, but also observed a mixed effect on fluency (improved breakdown and repair fluency but less speed fluency). Hidalgo and García Mayo (2019) analyzed the impact of these two forms of task repetition on EFL children's oral output while engaged in a collaborative writing task. Contrary to most previous research, they found that the children in the procedural task repetition condition focused on form to a higher extent than those in the exact task repetition group.

Definitely, the issue of task repetition is an intriguing one which could be fruitfully explored in further research. Manchón (2014) bets heavily on the potential of task repetition, and suggests extending this technique to the written mode, since the majority of studies on task repetition have been approached from the oral mode. Additionally, children's language learning process remains relatively under-represented in studies on SLA. Thus, further work is required to deal with the effects of task repetition on children's collaborative writing tasks. The analysis of the children's written output could produce interesting findings that account more for the impact of these task repetition types on the complexity and accuracy of the texts, and that determine

children's learning and retention of the knowledge they co-construct during collaborative writing tasks (Hidalgo & García Mayo, 2019).

(e) Collaborative tasks

Last, but not least, another technique which has been found to be beneficial for drawing learners' attention to form is collaborative work (R. Ellis, 2016). Most of the research analyzing attention to form in peer interaction has been undertaken from a psycholinguistic perspective (Long, 1996). Swain (1998, 2000) argues, however, that scholars working from a sociocultural perspective found that collaboration could be a different way to attract learners' attention to form while maintaining a primary focus on meaning. It has been acknowledged that collaborative tasks encourage children to consciously reflect on their own language use during meaning-based lessons (Barnawi, 2010; Calzada & García Mayo, 2020a, 2020b; Hidalgo & García Mayo, 2019; Leeser, 2004; Storch, 2016, 2018). As will be further discussed below (see section 2.3.1), collaborative tasks facilitate the noticing of gaps and a focus on formal aspects of the language. They also provide learners with opportunities to scaffold each other, co-construct meaning and thus collaborate in the solution of their language-related problems in ways that can facilitate L2 learning.

To conclude this section, we have seen that the literature identifies several effectual types of techniques to attract learners' attention to form during L2 learning communicative tasks such as text enhancement, corrective feedback, pre-task planning/preparation, task repetition or collaborative tasks. This field of research, however, has raised many questions in need of further investigation, not only to shed light on the mixed results obtained, but also to know more about aspects and populations which have been out of the scope of SLA research for many decades: FonF in L2 writing and with YLs.

2.2.2.3 Operationalizing FonF: Language-related episodes

So far, we have briefly explained what FonF is, why it is so highly valued among SLA researchers, and how it can be prompted in the L2 classroom, but we still do not know

how FonF has been operationalized and measured in practice. The following is a rough definition and description of the units of analysis normally used by researchers to analyze and quantify FonF episodes.

Research has analyzed attention, a cognitive process happening within the learner, by taking segments of interaction in which a change of focus from meaning to form is immediately observable. From this approach, interaction is primarily conceived of as an ideal scenario for negotiation of meaning, modified output, and negative feedback. Nevertheless, attention and noticing can also take place beyond CF and negotiation of meaning episodes (Fernández Dobao, 2014). Following Swain (1998, 2000), sociocultural researchers have employed a different, more encompassing unit of analysis to capture occurrences of attention to language form: the *language-related episode* (LRE). An LRE has been referred to as '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, 1995, p. 326). They are incidental and determined by learner needs and interests (Williams, 1999).

LREs can thus include segments of interaction in which learners may a) question the meaning of a linguistic item; b) question the correctness of the spelling or pronunciation of a word; c) question the correctness of a grammatical form; or d) implicitly or explicitly correct their own or another's usage of a word, form or structure (Leeser, 2004). As observed by Williams (1999), LREs may entail multiple interactional moves such as explicit and/or implicit feedback, negotiation of form and meaning, requests for assistance and metalinguistic comments, in the TL or in the students' L1, all evidence that learners are paying attention to language and that they are aware of the gaps between their own or their peer's IL and the TG.

The following example illustrates an instance in which Student 1 produces a lexical LRE by questioning the meaning of a Spanish word. Student 2 answers the question correctly:

(1) Student 1: ¿Qué es . . . el principio? [What is . . . 'el principio'?]
Student 2: Ummm . . . .like empezar. [Ummm . . . like to begin.]
Student 1: Hmmm?

Student 2: empezar . . . el principio . . . the beginning.
[to begin . . . the beginning . . . 'the beginning']

(Leeser, 2004)

Example (2) is an instance of a grammatical LREs including explicit corrective feedback. Student 2 corrects the pronoun used by Student 1, and provides the reason for his/her correction:

(2) Student 1: She–
Student 2: (laughs) He-He
Student 1: Is wearing–
Student 2: He. It's a boy. Is he
Student 1: Oh! She wearing
Student 2: He (laughs)
Student 1: Ha! He wearing

(Collins & White, 2019)

LREs are usually classified according to the linguistic category that participants concentrate on, but research has also analyzed the quantity and quality of the LREs produced by the students while performing collaborative tasks. For instance, some researchers have looked at the use and type of metalanguage (e.g., Fortune, 2005); others have paid attention to the grammatical or lexical nature of the linguistic items as well as to the resolution of the episode (e.g., Leeser, 2004; Williams, 2001). Others have examined the procedures used by the learners to solve the problems they posed to themselves (e.g., Cánovas Guirao, 2017; Coyle & Roca de Larios, 2020); and some others have classified them depending on the complexity and directness of the LRE along with self-repair (e.g., Adams & Ross-Feldman, 2008). On the other hand, Swain (1998) cautions that some factors may influence these characteristics, in such a way that what may be captured by a group of learners may not be the case with another group. These factors include age, level of proficiency, task features, background or educational setting, and roles and relationships assumed by students during paired/groups interaction, to name but a few (Collins & White, 2019).

Again, Collins and White (2019) observe that two territories are yet to be uncharted in the LREs domain. One dimension that needs further research is students' spontaneous attention to language while engaged in communicative tasks with their classmates as part of their regular routine. The general tendency is for researchers to introduce these tasks (e.g., Kim & McDonough, 2011; Swain & Lapkin, 2001), and while they may have pedagogical value, they are not necessarily activities that the teacher generally uses with their own students on a daily basis. Another under-researched activity that Collins and White (2019) bring into focus is the analysis of LREs produced in collaborative writing. Collaborative writing tasks such as dictogloss[1], jigsaw, text reconstruction, and composition tasks are tasks which students are asked to perform in pairs or groups to produce one final jointly written text (Swain, 2001). As mentioned earlier, these types of tasks are one way in which attention can be attracted toward form during communicative lessons. The need to agree on what to say and how to say it leads students to question their language use and to work together in the solution of their linguistic problems, thus producing LREs. Research suggests that, by triggering LREs, collaborative tasks foster L2 learning (Cánovas Guirao, 2017; Fernandez Dobao, 2014; Storch, 2001; Yang & Zhang, 2010). On the other hand, studies comparing FonF across these collaborative writing task types have revealed that tasks which have a more structured design, such as text reconstruction and cloze tasks, tend to prompt a higher number of LREs and specially more grammatical LREs than dictogloss, jigsaw, or composition tasks (Storch, 2001; Swain & Lapkin, 2001), from which it can be deduced that not all collaborative writing tasks seem to facilitate learning in the same way.

Although we provided earlier a definition for LRE, for the purpose of this dissertation we consider more convenient to adopt a more accurate definition by Qi and Lapkin (2001, p. 287) which helps understand more clearly what an LRE is in those studies addressing both writing and comparing. In this way, LREs can be applied to the comparison stage as well. Therefore, an LRE in our study refers to:

[1] Dictogloss is a classroom dictation activity, mostly used to teach grammatical structures, in which learners are required to reconstruct a short text in pairs or small groups by listening and writing down key words (Wajnryb, 1990).

> a segment of the protocol in which a learner noticed a language-related problem he/she encountered while comparing his/her text to a model and addressed it either by accepting the model and providing a reason, or only noticing the difference without giving a reason.

As can be seen, this definition encompasses not only the general meaning of an LRE, but also how these episodes can be captured during collaborative writing tasks. That is, it also takes into account those instances of language-related noticing which are not verbalized by the learners (e.g., self-correction), but are captured by the teacher, researcher or in a videotape as LREs. This issue is important, and even more so when working with children, because there are times when learners think to themselves and make corrections on their shared written compositions which are never spelled out but are still language-related problems. In many cases, there is a fine line for researchers between what learners perceive as a language problem and what they simply think to themselves in terms of meaning-making. In this sense, this definition captures to perfection this subtle difference.

Another issue which has not received much attention is the use of LREs by primary school children learning a FL. Attention to form during pair/group tasks has been examined mainly with adults (e.g., Leeser, 2004; Williams, 1999) or adolescents (e.g., Basterrechea & Leeser, 2019; García Mayo & Basterrechea, 2013; Swain, 1998; Swain & Lapkin, 2001), who are usually more language-focused than children (Collins & White, 2019). However, there is some evidence that children, when guided, are also able to attend to language while expressing meaning (e.g., Bouffard & Sarkar, 2008; Calzada & García Mayo, 2020a, 2020b; Harley, 1998; Horst, White & Bell, 2010, as cited in Collins & White, 2019). Given the growing number of schools introducing FLs at an early age, more research with this population is needed to look into the situations in which children spontaneously talk about language while engaged in pair and group work tasks designed by their teacher as part of the classroom curriculum (Collins & White, 2019). How do they proceed when they come across a language problem? How do they interact among one another to tackle those difficulties? What linguistic features are they usually concerned about?

To sum up, this section on FonF has reviewed a number of aspects which are key to understanding the facilitative role of both (collaborative) writing tasks and WCF as regards attention and noticing. The FonF pedagogical intervention was born as an alternative 'solution' to exclusive and single-minded approaches, namely the FonFs and the focus on meaning paradigms, by creating a symbiosis of such practices which would allow students to focus their attention on linguistic aspects while they are engaged in meaning making. However, the concept is still 'under construction' in the sense that both researchers and teachers are yet in search of effective and appropriate ways to attract learners' attention to form without isolating these forms from their meaningful context. Text enhancement, CF, pre-task preparation, task repetition and collaborative work are some of the techniques that have been proposed and researched in different contexts and with different cohorts of learners, and so has the operationalization of LREs. Interestingly, collaborative writing and WCF have a place in all these areas. What is more, the need to know more about the existing procedures for attracting attention to form and to look for new ones is also awaking interest in their application on the written mode among SLA researchers. Consequently, it comes as no surprise that there are increasingly more and more calls for studies on FonF applied to collaborative writing and/or WCF. Similarly, as new fields of inquiry emerge, so does the interest in how children would behave in these same scenarios. Although it is true that these calls are gradually being answered, research to date does not allow a clear answer yet. Therefore, these two under-explored areas together give us one of the rationales for exploring the characteristics of the LREs in primary school children as they are engaged in a collaborative writing task.

So far, we have explained the value of WCF from interactionist- and FonF-oriented perspectives. In the next section, we will introduce and explain the next cognitive theory to follow the discussion on what the language learning potential might be for WCF to play a role in L2 learning and acquisition.

2.2.3 Skill Learning Theory

Skill Learning Theory (DeKeyser, 2007) argues that language learning is not different from any other learning of skills, as they both involve cognitive processes that result in complex behavior as a consequence of the mastery of simple processes. The theory

states that there is a general learning mechanism responsible for the acquisition of new skills through a three-stage procedure: First, declarative knowledge (or explicit knowledge) will be obtained (i.e., grammar rules), then this knowledge will be proceduralized (become implicit), and finally it will be automatized. Therefore, students start learning something through explicit processes, and with practice, move into implicit processes. More specifically, the automatization of a language 'skill' requires time and practice, and this is the role that Skill Learning Theory researchers give to (W)CF: a learning opportunity for automatizing knowledge (Bitchener, 2012; Wigglesworth & Storch, 2012). According to Bitchener (2012), WCF is one of the ways in which explicit knowledge can be gained and automatized through prolonged systematic practice.

Central to skill learning theory is the *Transfer Appropriate Processing* principle, which basically claims that the characteristics of the context where knowledge is acquired should resemble the most the characteristics of the context where that knowledge will be retrieved (Segalowitz & Lightbown, 1999). Accordingly, L2 knowledge that is acquired in a communicative task will be beneficial when students are engaged in a communicative situation outside the classroom. Learners, thus, need to attend to meaning as a means to develop their communicative skill. From here one could expect then that WCF which is implemented in communicative contexts will allow learners to retrieve knowledge and put it into practice, thus contributing to the automatization of linguistic forms. Indeed, scholars working in the Skill Learning Theory framework have found WCF to be advantageous for the automatization of explicit knowledge through communication practice. Both the conditions of the teaching context and learners' orientation (to meaning or form) are believed to have an impact on the power of CF (Sheen, 2004).

Thus far, the present thesis has attempted to explain the rationale for expecting WCF to be beneficial for L2 learning from a cognitive lens. The aforementioned concepts and hypotheses such as the Interaction hypothesis (Long, 1996), the focus-on-form paradigm (Long, 2000), or the Skill Learning Theory (DeKeyser, 1997) encompass constructs that complement one another in trying to explain the L2 acquisition process. Since recent studies have shown that it is through writing and the provision of WCF that students activate such cognitive processes, interest in the potential of writing has come

to the fore. What is more, some authors have favored writing over speaking in relation to language use since, given the off-line nature of writing, students have more time to pay attention to form, to draw upon their explicit knowledge and to consider any feedback they may receive, making way for all the processes that are otherwise hampered in oral tasks. Therefore, the reasons behind the potential of writing are partly rooted on the idea that learning happens precisely due to the nature of the writing process. Besides, notwithstanding the increasing interest in the analysis of early language learning, studies on children's L2 writing and feedback processing in school contexts remains quite an uncharted territory within the field of SLA. As progress grows in each of the mentioned cognitive areas, more studies on writing and younger populations are needed.

However, there are different lenses through which the potential of WCF can be explained. Now, we will attempt to move the construct of WCF from a concept that is regarded as more static in the cognitive framework to a phenomenon which changes depending on numerous factors, like the context where it takes place. In the next section, we will follow this line of research through a lens that contemplates WCF within a context and not as an isolated procedure.

2.3. The sociocultural perspective on WCF

As Fernández Dobao (2014) indicates, research couched within a cognitive dimension attempts to examine the relationship between tasks and the occurrences of negotiation of meaning between learners and to gauge the effects of task features on effective language use and processing. Interaction between learners is believed to bring about L2 learning, as long as input is comprehensible (Long, 1983), negative feedback (either explicit or implicit) is provided and modified output occurs in L2 interaction (Long, 1996). In addition, Swain and colleagues (Swain & Lapkin, 1995, 2001) have extended the functions of output and have also admitted the importance of collaborative dialogue that occurs in learner interaction. Actually, Swain (2006) maintains that collaborative dialogue is a crucial source of learning, because it is the moment when students make meaning of their interaction and also shape their knowledge.

Swain's (2006) inclination toward collaborative dialogue brings us to Sociocultural Theory (SCT), which offers a different point of view on the role of interaction in L2 learning and is remarkable for the types of insights it proposes about the learning process, in conjunction with how students respond to and use (or not) the CF they are given (Bitchener, 2012). Having its genesis in the work of Vygotsky (1978), SCT claims that all human cognitive development (including language development) happens in interaction (not as a result of interaction) between individuals and learning results from the internalization of these interactions.

For L2 novices, this is expected to occur particularly when they are provided with opportunities to collaborate with more competent L2 speakers such as teachers or more knowledgeable peers (Bitchener, 2012). Nevertheless, not all forms of assistance are useful and result in development. Vygotsky (1978) differentiated two levels of cognitive development: the actual level of a novice or a representation of what the learner can do/knows independently; and the learning potential of the learner as realized through assistance of an expert. The distance between these two levels, which cannot be too big if we want to assure development, is encapsulated in the construct *Zone of Proximal Development (ZPD)*, technically defined as 'the distance between the actual developmental level as determined by independent problem solving and the level of potential development as determined through problem solving under adult guidance or in collaboration with more capable peers' (Vygotsky, 1978, p. 861). Therefore, according to SCT, development can only occur if the assistance given bears in mind the ZPD. The form of assistance adjusted to a learner's potential level of development is known in the literature as *scaffolding* (Storch, 2019).

From these positions, Storch (2018) outlines three broad ways of understanding WCF. One of them is regarding WCF as one form of assistance to fill the learners' gaps, and its effectiveness will depend on whether it represents tailored assistance within the learner's ZPD (Storch, 2018). By filling these gaps with the adequate help, the ZPD is constructed. Apart from the constructs of the ZPD and scaffolding to evaluate the nature and appropriateness of the feedback received, another key concept to understand CF is *mediating tools*. Vygotsky's main contention is that cognitive development is mediated by culturally constructed means or tools. Lantolf and Thorne (2006, p. 79), define those

tools as '[...] the process through which humans deploy culturally constructed artifacts, concepts and activities to regulate (i.e., gain voluntary control over and transform) the material world or their own and each other's social and mental activity'. Language is believed to be a symbolic tool, and thus one of those regulatory means which gives the expert and the novice the means to interact. Therefore, L2 learning is overall seen as the ability to gain voluntary control over the TL as a mediational tool. In the case of WCF, tools can have an effect on the provision and processing of feedback. To understand this, Storch (2018) sets an example of WCF delivered in the form of marginal comments. By using this feedback type, we may observe an effect on the quantity and the nature of the comments written by the teacher and also on the way learners engage, process and appropriate the feedback. Consequently, the mode of feedback provision used will have a particular impact on the learners' engagement and response to the WCF and, subsequently, on language development.

One of the criticisms that studies on WCF have received is that they are usually decontextualized, introducing feedback as a linear activity: the teacher provides feedback and the learner responds to that feedback (Goldstein 2001, 2006, as cited in Storch, 2018). No information is given about the context, or teachers and learners' goals and beliefs. The last key construct described by Storch (2018) is the one that views WCF as an activity and analyzes context-specific and individual factors that exert an influence on the provision of and response to WCF. Developed by Leont'ev (1978, 1981, as cited in Storch, 2018), *Activity Theory* is a model which can be used to observe and explain human behavior in general, and to understand why CF may or may not be responded to in writing tasks in particular. Its main premise is that human actions can be explained by reference to their objectives, emotions and context in which the activity takes place. For example, in the case of WCF, Activity Theory can be used to explain why some learners may focus on accuracy while others may focus on fluency or are less inclined to attend to and respond to feedback. In accordance, individual and context-related variables, alongside the nature of the WCF, will allow us to transfer findings from one context to other L2 learning contexts (Storch, 2018).

Based on these concepts, WCF is thus understood from a SCT perspective as a socially mediated form of assistance on errors which is attuned to a learner's potential level of development and also considers the contextual and individual factors as well as

learners' responses to the means used to provide WCF. As expected, the sociocultural approach also encourages the use of collaborative tasks to co-construct new knowledge. More specifically, SCT claims that through collaborative tasks teachers and learners jointly work in the assemblage of the learner's individual ZPDs and subsequent L2 learning. The ultimate goal of this process is learner's self-regulation (i.e. being able to use the L2 independently and autonomously) which can be achieved by using CF as a mediator (Lantolf & Aljaafreh, 1995). Therefore, the potential effects of WCF could be boosted if combined with collaborative writing.

2.3.1 Collaborative writing

Writing has generally been conceived of as a solitary activity. However, over the years, interest in collaborative writing tasks has increased among researchers and educators due to the language learning benefits this practice offers (Storch, 2019). As previously defined, collaborative writing is an activity that requires two or more learners to be involved in all the writing process to jointly produce a single text (Storch, 2002, 2019, 2021). In composing the shared text, the co-authors need to negotiate and agree about which ideas to include, how to organize them, and how to best put them into words. Therefore, the combination of both oral communication and writing throughout the text composing process is believed to offer not only the advantages provided by writing tasks (its off-line nature), but also the advantages provided by speaking tasks (the presence of an audience and the provision of immediate feedback). That is, writing in collaboration offers more benefits than writing or speaking tasks alone (Storch, 2021).

The act of using language to discuss about how to solve language problems is referred to by Swain (2000, 2006) as *languaging*, and occurrences of languaging have been operationalized as LREs. In the case of writing as a solitary activity, languaging is self-directed (private speech), but when learners write collaboratively (collaborative dialogue), languaging refers to the social construction of meaning through deliberations about linguistic problems which usually entail metatalk of aspects of the language (Wigglesworth & Storch, 2012). Swain (2006) maintains that both forms of languaging can lead to acquisition of new knowledge or consolidation of existing one, that is, languaging is a source of learning. Storch (2021) rightly observes that in collaborative writing tasks, both forms of language are present. In presence of others, private speech

is usually externalized, in such a way that it is audible and can elicit a response from their peer writer. Thus, private and joint deliberations during collaborative writing tasks can provide learners with opportunities to pool their linguistic resources, reflect on and deliberate over alternatives, co-construct knowledge, and give each other immediate feedback. As a result, students may strengthen their awareness of the relationship between meaning, form, and function and extend their current linguistic knowledge toward their potential developmental level (López-Serrano, Roca de Larios & Manchón, 2019). This is why collaborative writing has been (and is still being) researched to analyze how social interaction contributes to language learning and feedback.

2.3.1.1 Collaborative processing of feedback

Collaborative processing of feedback is another issue that merits further attention. As the name suggests, collaborative writing involves merely the act of writing, but writing in L2 classes brings about feedback on that written output, be it individual or collaborative. Thus, in joint compositions, feedback is also (or should be) usually confronted in collaboration although the processing is individual. That said, after peers receive feedback from their teachers, they are usually asked to respond to the feedback together. What research has shown is that when learners jointly notice or identify problems in their drafts, they display different noticing strategies and provide each other with suggestions and counter suggestions (Storch & Wigglesworth, 2010a, 2010b; Swain & Lapkin, 2002). For example, it might be the case that one student is good at identifying problems related to form (grammar or lexis), while another student is good at recognizing problems in content, cohesion or coherence. Collaboration offers learners mutual support when it comes to noticing problems. In addition, if learners interpret problems in a different way, they will negotiate them by justifying with arguments, which helps them develop critical thinking in noticing their gaps in their compositions (Barnawi, 2010).

Research on collaborative processing of CF has also shown that the quality of feedback processing may be influenced by a number of aspects such as linguistic and affective factors (Storch & Wigglesworth, 2010a, 2010b; Swain & Lapkin, 2002), learners' goals (Storch & Wigglesworth, 2010b) and the form in which feedback is provided

(Wigglesworth & Storch, 2012). For example, regarding feedback type, Wigglesworth and Storch (2012) found that indirect feedback (editing symbols) generated greater engagement, whereas direct feedback (reformulations) resulted in more enduring language learning. The examples below, taken from Wigglesworth and Storch (2012), illustrate how two students respond to these two types of feedback. Example 3 shows how they engage with the reformulation provided. The corrections were: the modal verb ('would' instead of 'may'), the article ('a' instead of 'the'), plurality ('students' and not 'student'), and the placement of the adjective 'real'. The excerpt shows that learners just notice the feedback, but do not engage sufficiently with the feedback received. Student 1 shows surprise and Student 2 just reads the reformulation.

(3) Learners' engagement with reformulation

Original: In contrast, it is well known that exams would not reflect the real students performance.

Reformulation: In contrast, it is well known that exams may not reflect a student's real performance.

Pair talk:

Student 1: It is well known that the exam may not

Student 2: It is well known that exam

Student 1: reflect a student's real, ah? Students always perform

Student 2: a real performance

Student 2: Students always perform. . .

In example 4, three errors were indicated by the editing symbols placed above the underlined errors: preposition, word form and article. We can observe greater engagement, but only with the feedback on word form and choice of preposition.

(4) Learners' engagement with editing symbols

Edited version: Some people think that exams are very important

C F C

to acknowledge the students' understanding about the subjects.

Pair talk:

Student 1: ‘c’. . .word choice problem, change the word. Article problem. Right? ‘for acknowledgement

Student 2: for. . .acknowledgement. . .I think for. . .

Student 1: in

Student 2: important . . . important

Student 1: that exams are very important in

Student: 2: in

Student 1: in

Student 2: in acknowledging

Student 1: yeah. . . Oh yeah yeah in acknowledging

In this excerpt, students deliberate about which preposition should be used instead of ‘to’, and also about the word form ‘acknowledge’. Student 1 provides the correct preposition (in) and Student 2 comes up with the correct noun form (acknowledging). We thus see clear evidence of collective scaffolding: the students resort to their existing knowledge and jointly find a solution that they might not have found had they been writing on their own. Although the authors assumed that more engagement would result in a deeper processing, that was not the case, since direct feedback resulted in greater accuracy in the following drafts.

2.3.1.2 Patterns of interaction

When discussing the language learning potential of collaborative writing, we must also consider the caveat that not all learners are willingly involved in these joint writing tasks. This may have an indirect effect on the frequency and quality of languaging and, subsequently, on learning outcomes (Storch, 2016). These differences emerge because the patterns of interaction vary according to a number of factors which can be categorized as learner-related (L2 proficiency, personality types, experience, attitudes and goals) and task-related (the design and implementation of collaborative writing tasks) (Storch, 2016; Wigglesworth & Storch, 2012).

Storch (2002) was one of the pioneering studies to address the dynamics of working in pairs by analyzing the nature of dyadic interaction in a university-level ESL class. She distinguished between four types of relationships according to the level of learners' contributions to the task (equality) and engagement with each other's contributions (mutuality) and found that some patterns were more conductive to learning than others. When both equality and mutuality are high, the relationship is labelled *collaborative*. That is, pairs work together to solve a problem, showing collective scaffolding and high contribution and engagement with each other's suggestions. The second pattern is labelled *dominant/dominant* or *cooperative pattern*, because equality is high, but mutuality is low. The members of the pair try to dominate the interaction and do not engage with each other's contribution. Nevertheless, Storch (2013) further clarifies that, contrary to a dominant/dominant pair who shows discord and conflict, learners of a cooperative pattern attend only to their own contribution without manifesting much conflict. A case of low equality and mutuality constitute the third pattern, and implies that one partner is dominant, while the other is passive. Hence the pattern is named *dominant/passive*. The last type is the *expert/novice* pair, where we find low equality and high mutuality. In such pairs, one member takes the leading role, but elicit contributions from the other member. Storch (2002) found that collaborative and expert/novice pairs were the patterns that resulted in more learning as compared to dominant/dominant and dominant/passive pairs.

2.3.1.3 Research on collaborative writing

The issue of how different variables (task type, L2 proficiency, social dynamics or the outcomes of collaborative dialogue on L2 learning and on students' written output) can have an effect on learning is an important one and, as such, a considerable amount of literature has been published into peer collaboration. By way of illustration, the relationships that learners form have been reported in a number of studies in different instructional settings (ESL vs. EFL), and with different populations. For example, regarding L2 proficiency, Storch and Aldosari (2013) conducted a study with 30 EFL students in Saudi Arabia and found that pairs made up of students with similar L2 proficiency (high-high and low-low) were more involved in the collaborative tasks than mixed proficiency pairs (high-low). In mixed proficiency pairs, the low proficiency

learner's input was scarce, contributing very little to the decision making, producing little incidence of languaging, and thus establishing a dominant/passive relationship. The study also revealed that the highest number of LREs was found in the data of high-high pairs. The researchers deduced that in order to determine how best to assign students to collaborative writing tasks, it is important to take into consideration not only the L2 proficiency, but also the type of relationship learners form. In the same vein, Tedick and Young (2016) carried out research with younger learners in a Spanish/English immersion program in the USA. They found that proficiency-matched pairs working jointly on reading and writing tasks showed a balanced contribution, whereas in mixed proficiency pairs, the more proficient learners took an authoritarian stance while their counterparts adopted a more subservient role.

The outcome of collaborative writing tasks is another aspect which has also arisen interest among researchers. In a study on accuracy, Storch (2005) compared the quality of texts produced collaboratively and individually by intermediate/advanced university ESL students. Storch found that in collaboration the resulting texts were more accurate and of better quality. Follow-up interviews revealed that most students showed positive attitudes about the activity. Kim (2008) compared attention to language in the form of LREs and performance on vocabulary tests (pre- and post-tests) of 32 intermediate learners of Korean who completed a dictogloss task. 16 students performed the task in dyads and the other 16 did it individually. Although both pairs and individuals generated the same number of LREs, the students who completed the task in collaboration performed significantly better on the vocabulary tests.

On learners' use of L1, Storch and Wigglesworth (2003) looked into L1 use during a collaborative writing task with intermediate ESL adult learners who shared an L1. The study revealed that the students made very little use of their L1. From the interviews with the learners, the researchers found out that this limited use of the L1 had to do with the students' belief that in the ESL context they were expected to use the L2. Storch and Aldosari (2010) added two variables to the use of the L1. They addressed the influence of proficiency and task type on EFL students' L1 use. They also found a scarce use of the L1 by these learners, even in dyads made up of low proficiency learners. Regarding task type, language exercises seemed to elicit more L1 use than collaborative writing tasks.

Most of the studies that have examined collaborative writing have been undertaken with adult or adolescent L2 learners, whereas research conducted on the effect of children's oral interaction on their collaborative written production is less abundant (e.g., Calzada & García Mayo, 2020a, 2020b; Cánovas Guirao et al., 2015; Coyle & Roca de Larios, 2020; Imaz Agirre & García Mayo, 2019; Shak, 2006). For example, in the ESL context of Brunei Darussalam, Shak (2006) engaged 78 9-12-year-old children in a dictogloss task and addressed whether they found the task in tune with their interests, needs and motivation, and whether teachers found it suitable with regard to their learnability, teachability and task usefulness. The study reported fluctuations in children's attitudes towards the task and showed that teachers need to consider their children's L2 proficiency level, along with their (mental) age and grade level, before deciding to implement FonF instruction.

In a more recent study, this time in an EFL setting, Imaz Agirre and García Mayo (2019) analyzed the interactional patterns as well as the LREs of 32 dyads of 11–12-year-old Spanish EFL learners while they performed an oral task and an oral + written task. The learners were distributed into proficiency-paired, teacher-selected and self-selected groups. The results indicated that most pairs were collaborative in both task modalities, but the proficiency-paired group turned out to be significantly more cooperative. As for attention to form, the children produced more LREs in the oral + written task. Another study on collaborative writing with children is that by Calzada and García Mayo (2020a), who explored the impact of a dictogloss task on the 11-12-year-old children's development of the third person singular morpheme and articles. The children worked individually, in pairs and in small groups. The results revealed that pairs and groups equally focused on grammar and discussed other grammatical forms to a greater extent than the target ones. The authors also found no significant improvement in the posttest results for the target features and a slight advantage of pairs over groups and individuals. Also using the 3rd person singular morpheme, Calzada and García Mayo (2020b) analyzed the oral-written connection of low proficiency learners aged between 11 and 12 years old while completing a dictogloss task. The authors reported high levels of attention to form as opposed to meaning, but also more attention was allocated to other grammatical forms than the target *-s*. Concerning resolution, there were significantly more correctly resolved LREs (which were mostly incorporated into their subsequent writing) than incorrectly resolved or unresolved ones.

Regarding WCF, there are also studies that have examined how adult and adolescent learners collaborate in writing and benefit from different WCF techniques (e.g., García Mayo & Loidi Labandibar, 2017; Hanaoka, 2006; Martínez Esteban & Roca de Larios, 2010; Qi & Lapkin, 2001; Yang & Zhang, 2010). However, once again, research on WCF and collaborative writing with primary school children is basically non-existent in the EFL context (Cánovas Guirao et al., 2015; Coyle & Roca de Larios, 2014; Coyle & Roca de Larios, 2020; Coyle et al., 2018; García Hernández et al., 2017; Lázaro-Ibarrola, 2021; Luquin & García Mayo, 2020). As an example (see section 3.3.1), García Hernández et al. (2017) worked with 60 EFL children aged between 11 and 12 years old and examined the impact of reformulations on the learners' use of problem-solving strategies while they completed collaborative writing tasks. The results of the study showed that the children who had benefitted from this CF type seemed to have greater opportunities for language learning than those who merely repeated the writing task. In addition, high proficiency pairs used a higher rate of strategies after receiving the feedback than their less proficient classmates.

In conclusion, SCT gives grounds for the use of WCF as an L2 development facilitator, fundamentally by regarding feedback as a form of assistance provided by a more competent language user like a teacher or a more knowledgeable peer. Nonetheless, SCT argues that not all forms of assistance are equally fruitful. To be productive and drive L2 development forward, WCF needs to consider the student's current and potential levels of cognitive development (ZPD) and be dynamic. In other words, only WCF which is dynamically attuned to the changing needs of a learner leads to learning (scaffolding). In turn, language, as a symbolic tool, allows the expert to calibrate the feedback to the learners' needs. Language also allows the student to meditate about the feedback, and eventually to internalize and appropriate it. Other material tools such as WCF types could also determine how students engage with and respond to it.

As Bitchener and Storch (2016) nicely put it, apart from portraying the ideal type of WCF and the way tools mediate the delivery and processing of feedback, looking at WCF through the SCT and Activity Theory glasses can help answer the everlasting question of *why*. Analysis from an Activity Theory viewpoint can elucidate *why* educators adopt particular WCF practices, *why* learners respond differently to the

feedback provided, *why* WCF brings about L2 learning for some learners but not for others. As claimed by Activity Theory, student's engagement and retention of the feedback provided may be contingent not only upon the nature of the feedback, but also upon the individual and context-related factors surrounding the learner which determine their orientation toward the feedback activity: the objectives the student wants to achieve, including the cultural and social dimensions within the context in which the WCF practice occurs.

Finally, it is also important to remember that, in line with sociocultural approaches on the role of collaborative dialogue in SLA, and more specifically with Swain (2006, 2010), languaging generated during collaborative writing and feedback tasks is believed to provide students with opportunities to boost their knowledge of the L2. When dyads pool their L2 knowledge to compose a single text and then together discover the importance of the information offered by means of WCF, they engage in a series of cognitive actions and processes. SCT closes the theoretical framework supporting the potential of WCF. Now, we will move on to explain how all the cognitive and sociocultural theories introduced so far interplay during a single WCF episode and beyond.

2.4. Stages in the processing of WCF

As we have seen, the justification for the language learning potential of collaborative writing and WCF rests on a number of leading theoretical strands of SLA. These are the Interaction Hypothesis (Long, 1996), FonF (Long, 1996), Skill Learning Theory (DeKeyser, 2007) and SCT (Vygotsky, 1978). But, to what extent are these constructs pertinent to L2 learning in the written context? With these four general foci of SLA research in mind, we can now bring together the various strands (Figure 1), all of which inform the present research. In order to explain how the interplay between these constructs may have an effect on L2 learning, it is necessary to start by the language production on which feedback will be provided.

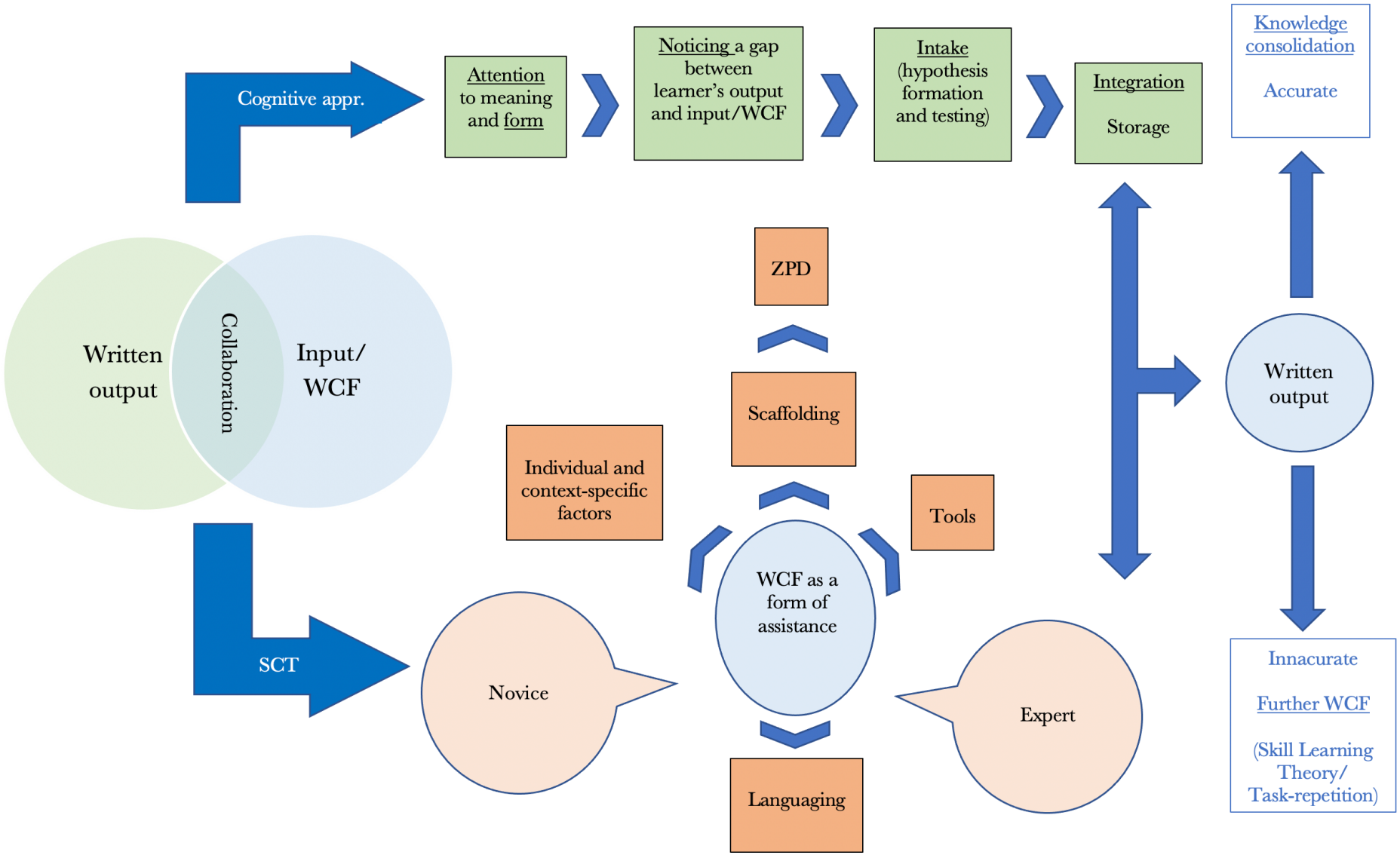


Figure 1. Stages in the processing of WCF from a cognitive and sociocultural perspective (adapted from Bitchener & Storch, 2016, p. 20)

Collaborative writing is output. Through our insights into the significance of the Output Hypothesis (Swain, 2000), we know that language production is the first step toward L2 learning. Although the Output Hypothesis usually refers to both speaking and writing, Swain frequently used collaborative writing tasks in her research to explore the nature of the dialogues that the activities elicited. We began this chapter by aligning ourselves with Manchón's (2011) notion of writing-to-learn. The arguments in favor of collaborative writing in L1 writing classes come from the learning-to-write dimension; while in L2 classrooms, collaborative writing is conceptualized as an activity that offers learners opportunities to learn the L2 (Storch, 2016). Thus, it would seem that producing language collaboratively in writing tasks is a vehicle which contributes to writing to learn the language. Collaborative writing tasks engage learners in deliberations about the TL (languaging) during the composition of a shared text. Once output is available for feedback, input and output processing come into play. In order to analyze the next step, we will make use of Bitchener and Storch's (2016) computational framework which is, in turn, an adaptation of Gass's (1997) integrated model of SLA. This framework describes the stages involved in the cognitive processing of a single WCF episode by which input becomes intake.

First, students obtain input in the form of positive or negative evidence on their written output. Then, attention is also necessary for students to uptake the feedback. Before they can benefit from WCF on their writing, they first need to notice a gap between their IL and L2 input. Whereas in oral contexts students are likely to overlook the feedback provided due to the fleeting nature of on-line interactions, it is less likely that this happens during a WCF episode because it is always explicit and because they have enough time to devote more attention to the feedback given (Bitchener, 2012). For this to occur, learners need to consciously attend to the feedback by (i) showing some degree of **alertness** or motivation and readiness to learn, (ii) **orientating** their attention toward linguistic accuracy and not only to meaning, and last (iii) **detecting** or cognitively registering the information about the L2 provided by the feedback. In doing so, the learner is ready to notice a mismatch between their output and the input provided in the form of WCF.

The next stage shows that input/WCF needs to be comprehended before it can become intake. Nevertheless, in Gass's (1997) view, comprehension is a gradual phenomenon

which may begin with the comprehension of meaning and progress to the comprehension of other components such as form and structure, which are essential for acquisition, and here is when FonF comes to the fore. As discussed above, FonF (Long, 2000) is a construct that supports the idea that effective L2 pedagogy should involve attention to linguistic form. One of the challenges among researchers and teachers, however, is how learner attention can be directed to linguistic forms during communicative tasks. We saw that, among the various types of FonF, both WCF and collaborative writing tasks constitute ideal contexts to bridge the gap between form and meaning. On the one hand, WCF represents a reactive FonF practice that takes place in a context of negotiation of meaning and form, and it is used to induce students' attention to form while completing a task in a personalized and individualized way. On the other hand, collaborative writing tasks can certainly facilitate noticing as students deliberate about language problems. What is more, collaborative writing and WCF activities lend themselves to the implementation of other ways in which learner's attention can be drawn to L2 features. For instance, through the use of models (see section 3.3.) as a WCF technique, textual input can be enhanced to highlight saliency of the L2 forms, causing noticing of the targeted L2 form. Another FonF trigger could be a pre-task session or some type of guided instruction on how to make the most out of the feedback delivered. Pre-task preparation intended for aiding learners in comprehending the nature of their errors could strengthen the quality of their noticing and processing of WCF. Finally, the mere repetition of collaborative writing tasks along with repeated exposure to WCF enables learners to focus their attention more closely on formal aspects. When they have to deal with the same or a similar task, different language dimensions will be attended to. Therefore, task repetition looks like a suitable context for FonF and also offers learners many opportunities for feedback.

The next stage is the internalization of comprehended input to become intake. This occurs when students match the input/WCF provided with their existing knowledge through processes of hypothesis formation and testing. When output is modified, learners will either confirm or reject their original hypotheses. Confirmation occurs when the input/WCF coincides with the learner's existing hypothesis and rejection when it does not match their hypotheses. In doing so, the learner has the opportunity to formulate a new hypothesis and have it confirmed or rejected next time they receive input.

Still, this processing of feedback may not be as straightforward as it seems. In a study carried out with high school EFL learners, Simard, Gunètte and Bergeron (2015) reported that although students understood the WCF for the most part, some feedback resulted in erroneous hypotheses about the intent of correction, especially with indirect WCF. They argue that if learners are not able to interpret the WCF correctly, reflecting on language effectively is almost impossible. In a similar vein, Qi and Lapkin (2001) suggest that for L2 learning to occur, students need to understand the nature of the gap between their IL and the TL, and therefore noticing without understanding cannot have the same effect on L2 learning as noticing with understanding. Likewise, Coyle and Roca de Larios (2014) observed that their young EFL participants misinterpreted elements that were noticed in the feedback but only partially understood, resulting in ill-formed language output. Consequently, it would seem sensible to think that utmost syntactic and semantic comprehension of WCF is a necessary condition for L2 development (Cánovas Guirao, 2017).

Notwithstanding the above, we know that hypothesis-testing in the written mode is advantageous in the sense that learners have enough time to retrieve existing knowledge from the long-term memory and compare it with the information offered in the feedback. Furthermore, it seems that learners feel more comfortable when processing feedback in this modality than when doing it in public, where issues of identity may be more vulnerable if hypotheses happen to be incorrect. As Williams (2012, p. 328) writes, 'the written mode can be a lower stakes arena in which to test out emergent forms'.

Written output produced in response to WCF input will be the evidence of whether or not the learner has initiated the process of developing new knowledge. If the output is not accurate, further exposure to feedback will be required for learning to move forward. On the other hand, it might be the case that the effect of WCF was delayed and was to come out in subsequent output. Complicating matters, however, Bitchener and Storch (2016) argue that the fact that the student's production of accurate output as a result of feedback processing is an indication that L2 development has started is not proof that new L2 knowledge has been consolidated, thus further practice may be needed to complete the consolidation and automatization process. This would explain why at times learners are able to produce accurate output, while other times they fail to

produce correct forms of the L2. In any case, what is relevant is that this inconsistency is an indicative of progress. In this sense, Ortega (2012) has managed to succinctly convey the essence of L2 progress by remarking that 'any SLA notion of progress has always been nonlinear, gradual, unevenly paced and often proceeding through interim nontargetlike (but developmentally helpful) solutions' (p. 408).

In summary, Gass's (1997) model claims that explicit knowledge (which can be acquired through WCF) is integrated into the student's IL through processes of noticing, comprehension, intake and integration. This notion clashes with the view of advocates of the non-interface viewpoint that WCF can only lead to explicit knowledge, whereas language production is only motivated by implicit knowledge. The theory that explicit knowledge cannot be transformed into implicit knowledge leads to the conclusion that learners' IL is unsusceptible to CF (Krashen, 1982). Consequently, CF is conceptualized as a useless practice which, at best, could bring about short-term improvements in accuracy, but not in L2 learning (Truscott, 1996). In contrast, theoretical arguments in favor of the interface view can be found in the Skill Learning Theory (DeKeyser, 2007) which argues that time and practice is required to reach the point at which explicit L2 knowledge is converted into consolidated and automatic L2 knowledge (implicit knowledge). Therefore, through sufficient and sustained practice of writing and further exposure to WCF on their written output, students can gradually automatize their explicit L2 knowledge to the point of making it suitable for acquisition (DeKeyser, 2007). In addition, research on collaborative writing (e.g., Storch, 2016, 2018; Wigglesworth & Storch, 2012) proposes that repeated collaborative writing tasks also supply learners with prolonged practice in discussing their language problems, in providing and obtaining feedback, and in rewriting. All these practices have the power to foster L2 learning. Nevertheless, more in-depth studies are needed to figure out the extent to which L2 acquisition occurs.

As discussed in section 2.3., SCT looks at WCF from a different angle to that adopted by the cognitive perspective. Whereas SCT regards WCF as fundamental for L2 learning, it rather concentrates on explaining why WCF may or may not be responded to in writing activities (Bitchener & Storch, 2016). What mainly differentiates sociocultural approaches from cognitive ones is that, in order to explain the whys of WCF, SCT does not look at how the brain works in terms of long-term memory,

hypothesis testing, etc., but rather puts a spotlight on the feedback delivered taking into account the individual learner's potential abilities (ZPD), how the feedback is processed as a form of problem-solving activity (languaging), the way it is provided (tools), and the individual and context-specific factors of the feedback as an activity.

From this perspective, WCF is seen as a socially mediated form of assistance delivered by a more competent speaker. Nevertheless, SCT also claims that not all forms of assistance are effective for L2 learning. Assistance needs to be dynamic and co-constructed between the participants during interaction so that it is calibrated to the learner's current and potential level of development (scaffolding), eventually leading to self-regulation. Therefore, applied to WCF, SCT argues that for WCF to be facilitative of L2 learning, it needs to be carefully adjusted to the extent that each learner is capable of understanding the feedback and profiting from it. Furthermore, feedback should encourage students to perform at their potential rather than at their current level of L2 capacities (Bitchener & Storch, 2016). This assistance makes it possible for the student to use knowledge internalized during the interaction in new situations.

All things considered, as opposed to Truscott's contentions concerning the harmful effects of CF, a large number of theories and studies have proved that WCF is beneficial for advancing L2 learning. Empirical evidence has shown that those learners who benefitted from WCF even on just one occasion promoted accuracy in new pieces of writing and this improvement was retained over different periods of time. For example, Ferris and Roberts (2001) worked with university ESL students and examined their abilities to edit their texts across three feedback conditions: errors marked with codes; errors underlined; and no feedback at all. They reported that the groups that had received feedback did significantly better than the control group (CG) who had received no feedback, and no significant differences were found between the feedback groups. More recent studies have analyzed the benefits of WCF by looking at the progress made in new pieces of writing. Bitchener, Young and Cameron (2005), for example, compared two types of feedback with a CG who did not benefit from WCF, and analyzed the extent to which the different feedback techniques had an impact on the accuracy of further pieces of writing. They found both types of WCF to exert a significant influence on certain accuracy-related aspects in subsequent writing, but not on others. These findings were confirmed by a series of experimental studies by

Bitchener (2008) and Bitchener and Knoch, (2008, 2009, 2010a, 2010b), who proved how diverse WCF types including direct correction and metalinguistic explanations resulted in improvements in students' written accuracy which were sustained over time. Along similar lines, R. Ellis, Sheen, Murakami and Takashima, (2008) compared the effects of focused and unfocused WCF on the English article system of Japanese university students and found that, irrespective of the feedback condition, both groups outperformed the CG, who received no correction. Sheen (2007, 2010) reported similar results in a study on the acquisition of articles carried out with adult intermediate ESL students of various L1 backgrounds. That is, the feedback groups performed better than the CG on the immediate post-test. In an effort to test some of Truscott's controversial arguments about the effectiveness of CF, Van Beuningen, De Jong and Kuiken (2012) explored the value of WCF with Dutch high school learners. Once again, they managed to provide further evidence in support of WCF by exhibiting the long-term effects of both direct and indirect unfocussed feedback on improved linguistic accuracy during text revisions and in new texts, as opposed to the CGs.

2.5. Conclusion

In this chapter, we aimed to support the role of WCF in L2 development drawing on theoretical approaches that have something to say about the cognitive and social processing of L2 information provided in the form of WCF. To this end, we began by explaining the difference between the notions of learning-to-write and writing-to-learn (Manchón, 2011), arguing that while the former sees feedback as a way of fostering students' competence in producing effective pieces of writing, the latter takes a different stance on CF, emphasizing its potential in aiding leaners' IL development. We also referred to a series of different arguments posed by CF contestants against the potential of feedback to help L2 learners acquire the TL. Afterwards, we provided an overview of the theories of SLA within the cognitive and sociocultural approaches and addressed to what extent they explain the language learning potential of WCF as well as why WCF may or may not result in successful L2 development.

The cognitive perspective centers on the mental processing of information about the TL (or of input in the form of either positive or negative evidence), and on its use in the accurate oral and written production. Within this approach, we explained the basic tenets of the Interaction Hypothesis (Long, 1996) which, in turn, comprises constructs such as input, output, noticing, feedback, attention, etc. We also described the FonF paradigm (Long, 2000) and contemplated writing as an ideal scenario for drawing students' attention to formal aspects of the language during communicative tasks. To finish, we described the Skill Learning Theory (DeKeyser, 1997), for which the automatization of a TL requires time and practice.

On the other hand, SCT (Vygotsky, 1978) offers a different stance on the role of interaction in L2 learning claiming that language development happens in interaction, and not as a result of interaction. From this perspective, WCF is seen as a socially mediated form of assistance provided by a more knowledgeable speaker. SCT does not just focus on WCF on collaborative tasks, but on feedback provided during human interaction in a purposeful activity and on explaining why WCF may or may not be effective. That is, SCT does not look at the mental processing of information just as cognitive approaches do, but rather looks at the feedback and how it is best delivered to the individual learners bearing in mind a series of factors. Within this perspective, we also talked about how collaborative writing tasks contribute to this social processing of input and we also brought into focus the importance that patterns of interaction may have on the frequency and quality of languaging, and subsequently on learning outcomes.

Finally, we closed the chapter by bringing together all these theories, constructs and hypotheses in an attempt to examine how they interplay during a single WCF episode. Succinctly, collaborative writing entails producing jointly discussed output enabling learners to solve problems and test hypotheses on which they receive feedback (Output Hypothesis). Through the provision of input in the form of positive or negative evidence, learners notice gaps (Noticing Hypothesis) in their IL and may focus their attention on formal aspects of the language (FonF) allowing them to formulate new hypotheses. With the subsequent production of output, learners put the new hypotheses to the test and the process starts again in such a way that the more opportunities students

have to engage in collaborative writing and WCF, the more consolidated and automatic L2 knowledge will be (Skill Learning Theory).

Furthermore, throughout the chapter we have hinted at the scarcity of studies undertaken on these issues with children. The fact is that the improvement in the ability of writing in a FL is hardly ever considered beyond the completion of textbook exercises. In conventional language classes, YLs usually copy, complete, choose, order, match, underline, circle, fill in gaps, and write short answers, but they are not often asked to write texts in English. Consequently, they are deprived of crucial occasions to test their hypotheses of the FL and to receive feedback on their written output. In this context, many EFL teachers appear to ignore that writing practice and WCF play a fundamental role in the linguistic development of their students. As a result, this practice does not usually have room in current pedagogical agendas. Therefore, given that there exist differences between adults and YLs, the discussion from both cognitive and social perspectives of how the latter differ from the former in terms of how they process and learn an L2 needs to be addressed in depth to contribute to the growing body of research on children's L2 learning processes (Cánovas Guirao, 2017). Giving YLs opportunities to engage in collaborative writing tasks and to receive feedback would, in consequence, seem paramount for putting all the above-mentioned theories to the test and finding out whether they also have the desired effect on children's L2 development.

In the following chapter, we will look at WCF in terms of the differential effectiveness of several methodologies on L2 learning. To this purpose, we will begin by describing the types of WCF that have been contemplated in the literature and implemented in the language classroom. We will then move on to point out the limitations of traditional forms of feedback and to introduce some novel or trendy techniques such as reformulations and models. To conclude the chapter, we will present an overview of research conducted on models with children. Thus far, the present dissertation has argued that there are cognitive and sociocultural theories that support the language learning potential of WCF but let us see now how WCF really works in practice.

CHAPTER 3: IMPLEMENTING WCF IN THE L2 CLASSROOM

The correction of errors is a common aspect of L2 teaching. For many teachers, correcting errors and commenting on learners' written assignments are part of their daily routine. However, the literature on CF that L2 learners receive in response to their errors is riddled with disagreement and conflicting results about its worthiness (Storch & Wigglesworth, 2010). Actually, ascertaining which types of (W)CF are more beneficial for L2 development and whether their degree of explicitness has an effect on this process are questions of both theoretical and pedagogical interest. From a theoretical perspective, these issues are decisive because if more explicit types of feedback are certainly more effective than less explicit types, theoretical considerations that justify and anticipate how students acquire the L2 need to take into account these differences as prerequisites for L2 learning. On the pedagogical front, they are also essential, as educators want to know how they can best help their students (Bitchener, 2012). Accordingly, the present chapter offers a review of the different types of WCF that can be found in the literature. The first part of the chapter provides a description of these feedback types as well as its main drawbacks. In the second part, alternative feedback techniques will be examined. Then, a discussion on the importance of models as a WCF type will be presented, and this will be followed by an analysis of current research on models conducted with YLs.

3.1. Types of WCF

There are different ways in which CF on L2 learners' writing can be provided. Methodologies of WCF may vary, for instance, as regards their focus, and their degree of explicitness including direct, indirect and metalinguistic feedback types. A number of arguments have been put forward in defense of these feedback methods and a wealth of studies has looked into their contentions. To begin with, the focus-unfocused debate is one of the greatest queries that researchers try to answer nowadays and has to do with whether or not learners' attention should be drawn to every error or to just one or a few categories at any one time (Bitchener, 2012; Van Beuningen, 2010). The unfocused approach entails the comprehensive correction of all errors in a learner's composition,

irrespective of their error nature. In focused (or selective) CF, on the other hand, the provision of CF is limited to a (number of) specific error type(s) and other errors are left uncorrected (Van Beuningen, 2010). Different claims have been made regarding the effectiveness of focused and unfocused CF. For instance, R. Ellis et al. (2008) argued that given that noticing and understanding are believed to be crucial for acquisition, focused CF must be more favorable to accuracy development than unfocused CF, as it is more probable that students will notice and understand feedback when they target specific language features. Against this backdrop, Bitchener (2008), Bitchener, Young and Cameron (2005) and Sheen (2007) claimed that an unfocused methodology may not be the best option since coping with a wide range of errors at the same time may produce a cognitive burden and a block of the feedback processing, even more so for L2 learners, who usually have a limited processing capacity.

Focused CF, however, has also been criticized for a number of reasons. According to Segalowitz (2000), with the implementation of this selective type of feedback, learners might find it more difficult to transfer what they have learned from the feedback to new writing situations. What is more, Ferris (2010) and Lalande (1982) observed that correcting specific language features could be insufficient and counterproductive in the sense that, when the teacher corrects their students' compositions, they generally aim at improving accuracy in general, and not at correcting the misuse of a particular form. In addition, perceiving that some errors have been addressed, while others have been left uncorrected might be rather confusing for learners.

The second dichotomy is that between direct and indirect CF. The main difference between both methodologies lies in that direct CF involves explicit indication of an error alongside the provision of the correct linguistic form, while indirect CF only signals that an error has been made. Indirect feedback can be delivered in a variety of different ways such as underline or coding of errors (Van Beuningen, 2010). Advocates of indirect feedback types claim that it encourages students to engage in a deeper processing of language and problem-solving by promoting the type of reflection that fosters long-term acquisition (Ferris, 1995; Lalande, 1982).

On the contrary, proponents of direct feedback such as EC (Chandler, 2003) argue that, unlike the indirect approach, direct CF offers learners sufficient information to solve

complex syntactical errors, thus reducing confusion. Chandler (2003) also observed that, while direct CF allows students to internalize the correct form, an indirect correction is not transparent enough for students to know whether their hypothesized corrections are actually accurate.

Offering students metalinguistic explanations in the form of comments, information, rules or examples of correct language use is also believed to foster learning, as the information given is even more explicit than direct CF. Nevertheless, it has also been put forward that the effectiveness of these methodologies can be contingent on a series of factors such as learner's proficiency, meta-linguistic awareness or processing capacity, the type of error targeted, what the teacher wants to achieve by providing feedback, or the nature of the information given in the feedback (Bitchener, 2012; Van Beuningen, 2010).

3.2. Limitations of traditional forms of feedback

As discussed above, learners usually find it difficult to simultaneously attend to both form and meaning in spoken language, but they can accomplish this in writing. Therefore, the written modality has the potential to afford ground for form focusing. Feedback on writing, then, should provide learners with opportunities to pay attention to form and help them make comparisons between their IL and the TL so that they may afterwards produce correct forms (Bitchener & Storch, 2016; Storch & Wigglesworth, 2010). Research on feedback has posed a series of limitations as to its effectiveness when it is delivered in the form of correction of errors. Although positive results have been reported regarding the effects of this feedback on a limited number of linguistic forms (e.g., the English article system or the simple past tense) (Bitchener, 2008; Sheen, 2007), some researchers have found fault with the merits of EC arguing that its implementation may (i) lack clarity, precision, and consistency (R. Ellis, 2009); (ii) lack sensitivity on the part of teachers to students' needs and ability levels (Hyland, 1998); (iii) produce confusion, anxiety and stress among learners given the amount of marks that makes it difficult to find out what is non-target-like (Hyland, 1998); (iv) result in passive reception and superficial hypotheses testing (Adams, 2003) and (v) lead to

superficial and transient memorization of the explicit rules or corrections (Truscott, 1996). Consequently, the value of traditional EC for L2 development is still a vexed question (Martínez Esteban & Roca de Larios, 2010; Polio & Williams, 2009).

Conversely, research has also analyzed whether leaving correction up to the student is possible and beneficial. Although it has been attested that self-correction is indeed beneficial for acquisition (e.g., Ferris, 2006; Lyster, 2004), several drawbacks have been found. For example, students can only self-correct if they have the necessary knowledge to do it. That is to say, according to Corder (1967), they can correct their 'mistakes' (made as a result of processing failures) but not their 'errors' (produced as a result of lack of knowledge). In addition, learners do not usually know what to correct, and this happens especially with YLs. They tend to have difficulty distinguishing between parts of speech and are unable to address any aspect other than grammar or lexis (Luquin & García Mayo, 2020). One solution to this problem could be the delivery of CF in a two-stage process: at Stage 1 students would have to edit their own texts, and at Stage 2 more explicit feedback from the teacher would be provided. One could argue, however, that such a practice is time-consuming and that it would probably be easier to simply offer an explicit correction (R. Ellis, 2009). Another issue concerning self-correction is that learners generally welcome feedback on their writing, particularly from their teacher (R. Ellis, 2009; I. Lee, 2004; Zhang, 1995).

Taken together, given that students need to notice the gap between their IL and the TL and test their hypotheses, along with the fact that the relationship between WCF and L2 development remains an open question, other feedback techniques should be deployed at least as partial options to traditional feedback methods. Two of the alternatives that have aroused special interest among researchers are reformulations (e.g., Adams, 2003; Qi & Lapkin, 2001; Yang & Zhang, 2010) and model texts (e.g., Hanaoka, 2006a, 2006b, 2007; Hanaoka & Izumi, 2012), since there is evidence that they both play an instrumental role in promoting noticing and metalinguistic awareness, and in engaging students in deeper processing than with more traditional feedback methods such as EC (Hanaoka, 2007a; Hanaoka & Izumi, 2012; Martínez Esteban & Roca de Larios, 2010; Qi & Lapkin, 2001). These more discursive feedback methods do not explicitly single out errors but treat the text as a whole providing appropriate language, organization,

mechanics, style and ideas for a given context, rather than offering lists of corrected errors, editing symbols or metalinguistic codes (Cánovas Guirao, 2017).

Reformulation is defined by Levenston (1978, as cited in Qi & Lapkin, 2001) as 'a native speaker's rewriting of an L2 learner's composition such that the content the learner provides in the original draft is maintained, but its awkwardness, rhetorical inadequacy, ambiguity, logical confusion, style, and so on as well as lexical inadequacy and grammatical errors are tidied up' (p. 281). By keeping the same content, the reconstructed text provides a native-like model so that the student can compare their own production with a native writer's version of it (Qi & Lapkin, 2001). According to this definition, reformulations fail to provide the writer with a wider array of alternatives above and beyond their own writing (Allwright et al., 1988; Coyle & Roca de Larios, 2014; Martínez Esteban & Roca de Larios, 2010; Yang & Zhang, 2010). In Allwright et al.'s (1988) own words 'a good reformulation may not be a good sample of native writing since it is limited by its faithfulness to the original writer's intentions (p. 254). Along these lines, Yang and Zhang (2010) maintain that reformulations appear to be useful at a sentence level, while models can provide students with a good sample for not only the specific sentence but also the whole discourse. What is more, from a practical viewpoint, the cost-effectiveness of reformulation makes it an infeasible tool for teachers to use in class. Models have therefore come to the fore to offset these shortcomings (Cánovas Guirao, 2017).

3.3. Model texts as a written feedback alternative

Modeling is another type of feedback technique which consists of providing learners with native or native-like texts that they compare with their original draft (Martínez Esteban & Roca de Larios, 2010). So, this is the most unfocused, indirect and least explicit type of feedback that can be offered to students, as learners' errors are not corrected nor flagged, but they have to strive to find out the implicit 'message' concealed in the model text (Murphy & Roca de Larios, 2010). Contrary to reformulations, models are written tailored to students' age, proficiency level, and also taking into account the content and the genre of the task, but not the original texts

written by the students (Cánovas Guirao, 2017; Martínez Esteban & Roca de Larios, 2010). What is more, although R. Ellis (2016) notes that explicit CF is more effective than implicit CF in that the former is more likely to assure attention to form, he argues that implicit forms of CF (such as models) have also proved to be beneficial and may even have a greater long-term effect.

Modeling is a relatively under-explored method for delivering feedback on learners' errors. As Johnson (1988) observed, traditionally, models were given to learners before getting them to write their compositions in an attempt to teach them how a text should be written in the TL. Nevertheless, this practice typically caused students to copy the model, which implied that they did not have opportunities to discuss about what the model offered in terms of language and content and, consequently, they were less likely to activate critical processes of attention and noticing (Cánovas Guirao, 2017). Nowadays, models are being rediscovered, since they are believed to play a more significant role than originally thought.

This type of feedback seems to be beneficial for guiding students' attention not only to a large amount of appropriate expressions, structures and lexis, but also to alternative ideas and content, an asset which is imperative for EFL learners as they are often deprived of opportunities to be in contact with native speakers (Manchón, 2009). Models also seem to encourage cognitive conflict, as they usually offer information that clashes with students' hypotheses about how language works, causing them to deliberate on alternative forms to convey their intended meanings (Cánovas Guirao et al., 2015; Coyle & Roca de Larios, 2020; Tocalli-Beller & Swain, 2005).

Because errors are not explicitly signaled, modeling has also proved to be useful for pushing learners to find their own errors, which could result in deeper processing (Sachs & Polio, 2007). Actually, the most important benefit of models is that they offer an ideal scenario for students to notice the similarities and differences between their IL and the TL, leading to a re-evaluation and confirmation of their L2 knowledge and thereon to modified output (Sachs & Polio, 2007). This advantage can be explained in light of Schmidt's (1990, 2001) Noticing Hypothesis and Swain's (1995) Output Hypothesis. Returning briefly to the theoretical framework, the Noticing Hypothesis (1990, 2001), on the one hand, claims that the more the learners notice, the more they learn. On the

other hand, the Output Hypothesis (1995) argues that output stimulates noticing. The procedure then would be the following: (i) when producing written output, learners may notice gaps between what they want to say and what they are able to say, and this is the moment when they gain consciousness of their own linguistic limitations; (ii) such noticing primes them to search the model text for a solution to their previous language limitations (Hanaoka, 2007; Hanaoka & Izumi, 2012; Izumi, 2003). In this sense, Johnson's (1988) claim that models should be provided right after students complete the writing assignment would lead learners to search the model for those features that turned out to be troublesome in their first writing, or even those which they decided to avoid due to a lack of resources (Hanaoka, 2006a), therefore fostering selective attention to particular linguistic features (Cánovas Guirao et al., 2015); and (iii) while reading the native-like model text, not only can learners notice gaps and solve their limitations, but they can also notice holes, because they are presented with alternative lexis, grammar, ideational and organizational input (Hanaoka, 2007; Hanaoka & Izumi, 2012).

In addition, as mentioned earlier, while reformulation is not the best option to use in large EFL classes, where teachers have little or no time to rewrite each and every one of their students' written texts, preparing only one model text for the whole class is much less time consuming (Ferris, 2010). These ideas imply that students will most likely regard models as the target-like versions of their drafts (Sachs & Polio, 2007) and view them as accessible (Lázaro-Ibarrola, 2009).

Taken together, models would seem to serve as a valuable feedback tool for EFL learners. Notwithstanding this great pedagogical appeal, there is a small body of literature that is concerned with the role of models in L2 acquisition.

3.3.1 Empirical research on models

There has been little research reported on the effectiveness of models as a feedback technique. So far models have been examined in both individual and collaborative writing with EFL high-school students of Spanish (Martínez Esteban & Roca de Larios, 2010; Montealegre Ramón, 2019), Basque-Spanish (García Mayo & Loidi Labandibar, 2017) and Korean (Kang, 2020); with primary school children (Cánovas Guirao et al.,

2015; Coyle & Roca de Larios, 2020; Coyle et al., 2018; Lázaro-Ibarrola, 2021; Luquin & García Mayo, 2020, 2021; Villarreal & Lázaro-Ibarrola (sumbitted)); and with Japanese ESL (Abe, 2008) and EFL (Hanaoka, 2006a, 2007) university students. Models have also been compared to reformulations with two Japanese EFL learners (Hanaoka, 2006b), and with Japanese (Hanaoka & Izumi, 2012) and Chinese EFL college students (Yang & Zhang, 2010). Likewise, the use of models has been contrasted with EC with primary school EFL learners in Spain (Coyle & Roca de Larios, 2014).

The typical sequence of research with models consists of engaging learners in a three-stage writing task involving (i) spontaneous noticing of linguistic problems while writing a picture-based story (Stage 1 (S1)), (ii) comparison of their initial drafts with the native-speaker model, which may offer solutions to their problems, and noticing (or not) of those solutions (Stage 2 (S2)), and (iii) rewriting of their original text, which works as an immediate post-test to see whether learners incorporate the solutions into the revised texts (Stage 3 (S3)). Some studies have included a delayed post-test which was carried out a week (Luquin & García Mayo, 2021) or two months later (Hanaoka, 2006a, 2007). When looking at the codification of noticing, the analytical category employed in these process-oriented studies usually include the LRE which is in turn classified into (i) the problematic features that students notice at Stage 1 (PFNs), the identification of a mismatch between the learner's written output and the model or the features noticed at Stage 2 (FNs), and the features noticed and incorporated at Stage 3 (FNIs). In practice, however, each study until now has addressed some or all of these stages in different ways and oriented their research toward particular aspects of students' writing and feedback processing. In what follows, the main findings and limitations of existing research with models will be summarized. After the summary, the reader will find Table 1 (p. 94), which displays the main features of each of the studies mentioned.

The results of studies on models generally point to a series of benefits for adolescents and adults. They are consensually considered to be valuable for encouraging learners to notice and incorporate mainly lexical features, over all in the case of higher proficiency learners (Abe, 2008; García Mayo & Loidi Labandíbar, 2017; Hanaoka, 2006b, 2007). In this regard, it has been attested that students usually notice lexical holes in their IL

when writing their initial drafts, and this leads them to a lexically-oriented search for solutions in the model texts delivered later (Hanaoka, 2007; Kang, 2020; Yang & Zhang, 2010). Models also seem to be beneficial for encouraging the noticing of alternative expressions and ideas (in particular for students working in collaboration) that learners use in their subsequent texts (Hanaoka, 2006b; Martínez & Roca de Larios, 2010). Research on the comparison between models and reformulations (Hanaoka, 2006b; Hanaoka & Izumi, 2012; Yang & Zhang, 2010) has highlighted different but complementary roles played by both feedback strategies. Yang and Zhang (2010) found that learners were able to notice most differences between their text and the reformulated one, but models provided them with a broader range of language input, particularly with new lexis and content as well as good examples of writing at the discourse level. In like fashion, Hanaoka and Izumi (2012) report that while reformulations allowed their students to find solutions to overt problems, models offered solutions for both overt and covert problems (features that are problematic during the written production, but that are not visible in the text). The authors come to the conclusion that reformulations are of help for students to distinguish linguistic errors, but do not provide alternative information that enables them to expand their linguistic repertoires. Models, on the contrary, have the potential to attend to form and meaning by expanding the propositional content of learners' writings (Hanaoka, 2006b). As for task completion, only Yang and Zhang's (2010) participants worked collaboratively during written production and feedback engagement. The authors found that having an opportunity to discuss with peers contributed to the improvement of the learners' subsequent texts. For their part, Martínez Esteban & Roca de Larios (2010) carried out a comparison between individual and collaborative writing. Results revealed that those students working collaboratively noticed and incorporated a larger number of features than their individual counterparts.

Although most of the research into models has been devoted to adolescents and adults, rather less attention has been paid to children on the use of models. Among the nine exceptions we find the work done in an EFL context by Coyle and Roca de Larios (2014), who analyzed the role played by EC and models on the noticing and subsequent revisions of 46 11- and 12-year-old Spanish EFL children engaged in a three-stage collaborative writing task. At the comparison stage, half of the pairs received their texts containing the teacher's EC, and the other half received two models. The authors

conducted a quantitative analysis of the LREs reported on the children's notes and their noticing was traced across the stages and into the revised texts. The children's texts were also analyzed to identify improvements in acceptability and comprehensibility. The results revealed that both forms of feedback mainly prompted the noticing and subsequent incorporation of lexical features. However, learners in the EC condition were found to notice more grammatical features at the comparison stage, which later resulted in higher acceptability and comprehensibility in the children's revisions. This led the authors to assume that due to the perceptual salience and personalized nature of EC, children did not have to resort to searching and interpretative procedures to identify differences between the feedback and their drafts, and so could devote their cognitive capacities to the perception and reporting of the differences spotted (Cánovas Guirao, 2017; Sachs & Polio, 2007; Santos, López Serrano & Manchón, 2010). The learners in the model condition, on the contrary, attended mostly to language chunks and content, which the higher proficiency learners could incorporate into their writing to a higher extent. The low proficiency children, however, used to simplify the chunks or recombine their elements to produce unique clauses formed by misinterpreting elements that had been noticed but had not been completely understood. Additionally, the authors point to the usefulness of collaborative writing as a site for leaning in the L2 classroom. Nevertheless, due to the lack of a CG, it is not clear whether the improvements in lexis observed in the children's revised texts was a direct consequence of the models delivered or of other variables. Also, in order to avoid misinterpretations or lack of understanding, the model should have been more carefully adjusted to the learners' needs. In any case, this study proves the merits of both feedback conditions confronted collaboratively for encouraging noticing among children and for helping them diversify their linguistic concerns.

In another study with young language learners at low proficiency levels, Cánovas Guirao et al. (2015) re-examined the effectiveness of modeling, this time including a CG in order to provide more robust evidence of the effect of models on children's noticing and writing performance. They worked with 20 10 and 11-year-old EFL children organized into ten sets of proficiency-matched pairs and randomly assigned to either a TG or a CG of five pairs each. The learners' pair talk and written notes produced during the composition-comparison-rewriting stages were analyzed to obtain information about the learners' noticing. Once again, the authors found model texts to

be helpful for drawing the children's attention toward lexis and chunks of language. They also observed that this feedback type fostered overall improvements in the quality of their writing, because YLs' tended to eliminate unnecessary content and produce more accurate descriptions of the picture prompt, as well as more coherently structured texts. Conversely, a lower number of formal features were incorporated into their subsequent drafts, so grammar was not the most attended aspect of the model text. In addition, and consistent with the literature on adults, findings showed that higher-proficiency learners noticed and incorporated more features than students with a lower proficiency. In view of the problems the children went through in the processing and incorporation of language features into their revised texts in syntactically acceptable ways, the researchers concluded that with younger lower proficiency learners, both the scope and quality of their noticing seemed to be essential for the fruitful use of model texts (Hanaoka, 2007). The limitations these YLs underwent during feedback processing as a consequence of their lack of L2 knowledge motivated the authors to stress the need to offer children guidance and practice in using models with the aim that they can better exploit the potential of this form of feedback. In their view, YLs need assistance with the extraction of ideas, lexis and form, as well as with how to allocate their attention in a more strategic way. Because of this, they propose that teaching children how to make the most out of models may improve the quality of their noticing. The main weakness with this study is that the above results should be considered as tentative given that only ten pairs were involved in the project. One of the drawbacks of conducting research with school learners is that sample sizes are usually too small to make generalizations, even more so when the reduced number of participants researchers usually have had to be cut in half to form pairs for collaborative work.

Later, based on Cánovas Guirao (2017), Coyle et al. (2018) undertook a longitudinal process-oriented study that also examined the role played by models in two multi-stage collaborative writing and feedback cycles over the span of five months apart, but this time including a period of instruction. The contribution of this study lies in that it does not account so much for the result, but for the process itself by identifying the sequential routes or trajectories the children might follow across tasks and keeping track of the changes. It is also unique in that the study adds a long-term dimension to previous one-shot studies, thus being the first to analyze whether the positive effects persist over time. Accordingly, 16 Spanish EFL 10 and 11-years-olds were divided into a TG, who

received instruction over a period of six weeks, and a CG, who did not. The LREs found in the collaborative dialogues and the written notes were analyzed across the three stages in both cycles, and the resulting children's texts were also examined to identify gains in the acceptability and comprehensibility of their written production. The authors then established connections between the students' noticing throughout the stages and its impact on their written output, thus identifying the trajectories followed by the children. In line with the afore-mentioned studies, the findings showed that models attracted most of the children's attention to lexis, although the teaching group was able to incorporate a wider range of language features than the non-teaching group. Results also indicated that YLs did follow a series of trajectories across writing and feedback processing tasks that entailed a broader spectrum of interwoven problem-solving strategies, noticing and uptake possibilities than originally identified in previous studies of WCF. The children's use of trajectories was found to exert a considerable influence upon their developing L2 knowledge. It was also evidenced that learners who had received instruction followed more beneficial trajectories than their peers. The study certainly opens a new dimension toward the processing of WCF but, as the authors themselves acknowledge, the small number of participants impedes the extrapolation of the results. Also, improvement on the part of the teaching group could also be attributed to a task-repetition effect and not only to prolonged training.

In a pilot study conducted to evaluate the feasibility of the present research project, Luquin and García Mayo (2020) reported different results. As usual in this type of studies, they analyzed what children noticed and incorporated in collaboration during a three-stage task. The participants were 12 11- and 12-year-old Spanish EFL students assigned into a TG, which received a model, and a CG, which self-corrected their texts. The children's dialogues and texts were examined to look into what language features the learners had noticed and subsequently incorporated. The unit of analysis was also the LRE, which was classified into different language categories. Contrary to previous findings, the results showed that at the writing stage, the children allocated more attentional resources to formal aspects. The authors attributed these results to the importance that grammar still has in elite schools of this type, but also to the fact that the best students were selected for completing the task. Actually, according to Leeser (2004), more proficient learners usually find it easier to process grammatical form better than less proficient learners, on the grounds that the former do not have to strive as

much with processing meaning during communicative tasks. The comparison stage, however, gave a new turn to the situation, as both groups focused mostly on content and lexical features, with statistically significant differences between groups. That is, the TG noticed significantly more lexical features than the CG. Significant differences were also found between Stage 1 and 3 concerning lexical LREs in the TG. In light of this change of focus, the authors assumed that the picture-based narrative task required students to describe a sequence of actions and, as a result, to concentrate on a specific set of lexical items. Again, these results should be seen as tentative given the small database. In addition, the fact that the twelve participants were the most proficient ones in their class does not seem to be very representative of the average learner. Finally, the study might have been more relevant if it had included a period of instruction by means of class discussion, for instance, or a delayed post-test.

Continuing with the process-oriented line of research, Coyle and Roca de Larios (2020) introduced a new variable, as they explored children's engagement with models in two different instructional settings: EFL and CLIL. To this purpose, 16 children from two different schools participated in the study. Four 10 to 11-year-old pairs from an EFL class, and four 9 to 10-year-old dyads from a CLIL class were engaged in a three-stage collaborative writing and feedback task. The children's pair talk, written texts and written notes/underlines at Stage 2 were analyzed across the three stages in order to establish connections between the children's original texts, the cognitive strategies used to process the feedback upon noticing a mismatch between the models and their own output, and the changes made in their subsequent output. The results revealed that both the CLIL and mainstream pairs focused mostly on lexis at the composing stage. At the comparison stage, while noticing surface-level differences between the model and their own text was the most used strategy by all pairs, only the CLIL pairs allocated their attention to new and alternative features in the feedback, which were fully or partially incorporated, thus improving their revised texts. Conversely, EFL children focused on confirming the similarities between their drafts and the feedback, which led them to repeat rather than improve their original writing. Finally, a great amount of the noticing from the models which was not reported in the pair talk also appeared in their revised texts. In view of these results, the authors conclude that children's previous learning experiences in CLIL or mainstream programs are likely to mold their cognitive responses to WCF. Moreover, they suggest that the students' low proficiency level may

have accounted for the limited noticing and the scarcity of metalinguistic reasoning. All the studies reviewed so far suffer from the fact that the number of participants is too limited to extrapolate the results, and it is also the case in this study. Moreover, the lack of a CG makes it difficult to discern whether the improvements observed in the learners' subsequent writing was a direct consequence of the feedback provided or of task-repetition effects. In addition, written notes were only asked at Stage 2, thus failing to complement the children's verbalizations at stages 1 and 3. Finally, although the authors claim that both groups are comparable, age and proficiency level are two variables that might have offset the advantages of one group over the other.

Some of the results obtained in the present study were recently published by Luquin and García Mayo (2021). This study aimed at analyzing the use of written models as a technique of WCF in two groups of young EFL learners who worked collaboratively on a picture story task. The participants were 38 11-12-year-old children divided into a TG, which received a model, and a CG, which self-edited their texts. Data were collected at three stages plus a fourth stage which served as a delayed post-test, and the authors analyzed LREs and CREs that occurred during writing, and more specifically those that had been noticed and incorporated. The results showed positive findings for the use of models as a way of WCF, as (i) the TG noticed a considerably higher number of features than the CG, (ii) the TG noticed significantly more lexical and content-related features than any other category and the difference was also significant with respect to the CG's noticing of the same features, while the self-correction group focused mainly on form and mechanics, and (iii) at Stage 3 and 4, the dyads in the TG incorporated significantly more features than their counterparts, most of which were discursive, mechanical and formal. This work included a delayed post-test, and a larger sample of participants in comparison to other research carried out with a similar population on models. In this sense, it contributed to the existing (and scarce) literature about models with more evidence of their benefits. However, the authors considered Stage 2 as the starting point, but failed to focus on the discussions that happened at Stage 1. That is, students created most of the episodes at Stage 1 and thus, it would seem necessary to discuss it. Therefore, a complete explanation where the linguistic modifications carried out at each one of the four stages are discussed and related is missing. In addition, the study focused on what YLs noticed and incorporated, but not on how accurately they

solved their linguistic and content-related problems throughout the stages. As for the general focus of attention, the authors considered that they failed to widen the scope toward other linguistic aspects different from lexis despite the training session, so they recommend more teacher-led sessions or even the use of different techniques to promote the noticing of a greater variety of features. Finally, the procedure of the first and last task was similar, so there might exist procedural task repetition effects that influenced the learners' performance.

The work of Lázaro-Ibarrola (2021) is based exclusively on model texts but focuses on three aspects about which very little is currently known: the impact of models on the accuracy of the incorporations and on the quality of the children's drafts as well as the potential differences between collaborative writing and individual writing, which was also examined by Martínez Esteban and Roca de Larios (2010). The participants in Lázaro-Ibarrola's study (2021) were 33 Spanish learners aged between 11 and 12 years old divided into a pair and an individual group, and the analysis comprised not only the amount and types of features, but also the correctness of incorporations and the comparison of the initial and final drafts using holistic and quantitative measures. The author found that, as in previous research, all the participants noticed and incorporated features, which were for the most part lexical. In addition, the students' second draft, when rated holistically, was significantly better than the first one. Yet, the quantity of noticing was rather low when compared to previous studies. As for the two groups, the results are substantially similar although, contrary to what Martínez and Roca de Larios (2010) found, individual writers noticed a higher number of features, although pairs were better able to incorporate features correctly. In her opinion, the reason why pairs did not notice as much as expected could be ascribed to the fact that they already solved some of their linguistic problems together, so they did not feel the need to write this noticing down. Conversely, individuals were deprived of the opportunity to deliberate and so tended to take more notes. The author proposes analyzing the pair talk to corroborate whether the noticing they discuss corresponds (or not) to the noticing they report in the notes. Concerning limitations, apart from the small sample size studies of this type usually have, one major drawback of the approach the author adopted is probably the use of 'all or nothing' criteria on the children's written production. By doing this, she overlooks other changes such as partial corrections, substitutions, new incorporations, etc. which would certainly provide us with more information on the

children's language development. Although the power of models to enhance noticing has been limited if compared with previous research, the use of model texts was undoubtedly beneficial for both groups. Consequently, in line with Luquin and García Mayo (2021), the author encourages teachers to use model texts in the classroom together with training sessions to facilitate a greater variety and quantity of noticed features and accurate incorporations.

The latest study on models is Villarreal and Lázaro-Ibarrola (submitted), which include a new component which is essential in YLs and could help us understand its influence on task effectiveness: motivation measures (see section 4.4). More precisely, this study explored the effect of models on the collaborative writing of children aged between 9 and 10 years old and addressed both draft quality measures and learners' motivation. To this end, 13 pairs of CLIL learners divided into a CG (who received no feedback) and a MG (who received feedback in the form of models) engaged in the typical three-stage task sequence of research with models, although there was no third session for the CG. In addition, at pre- and post-task, all dyads filled out an individual motivation thermometer in which they had to rate their motivation and select their reasons for their ratings in the thermometers. The children's written texts were analyzed both quantitatively through measures of complexity (grammatical and lexical features), accuracy and fluency, and qualitatively, following a 5-point scale rubric. The pre- and post-motivation scores were compared, and motives were analyzed qualitatively. About draft quality, the findings revealed a significant increase in complexity in the MG. As for the learners' motivation, the scores in both groups were mostly positive and evidenced a significant increase post-task, with collaboration being the most selected justification for their high rating. Motivation levels, nevertheless, differed considerably at the post-test: while the CG maintained them, the MG exhibited a significant drop. The main weakness with this study is that, as the authors themselves admitted, the study failed to further develop the reasons for the negative disposition of the children in the MG. This would also allow future investigations to test and refine them. On the other hand, a longitudinal follow up of the children would be interesting to observe how maturation influences learners' attention and motivation. Notwithstanding the above, this work certainly opens a new window into the array of individual learner-internal factors which mediate the way YLs engage in WCF processing, more specifically in models processing.

Table 1. Summary and Shortcomings of Studies on Models (adapted from Cánovas Guirao, 2017, p. 43-46)

Study	Feedback type and data collection method	Participants	Results	Shortcomings
Hanaoka (2006a)	• 2 models (picture prompt) • Note-taking	Individual: 37 Japanese EFL university students	• Models allowed L2 writers to notice solutions to covert problems (overall to partial avoidance types) that did not surface in the text. • Such noticing led the writers to incorporate those solutions into their revised texts.	• Note-taking falls short (some incorporations were reported). • The number of avoided features observed is too small to make any strong claims. • Incorporations coded as 'all or nothing' categories. • Lack of a CG to discard task repetition effects. • Lack of justification for the use of two models.
Hanaoka (2006b)	• Models and Reformulations (picture prompt) • Think aloud	Individual: one high and one low proficiency Japanese EFL student.	• Participants noticed mostly lexical features. • Models addressed alternative forms and developed the original content, while reformulations promote noticing of linguistic inadequacies. • Proficiency seemed to affect the number of FNs to a greater extent than the number of PFNs and the FNIs.	• The findings may not be extrapolated, as it is a case study. • Focus is only on the number of FNIs. • Failure to identify the source of the FNIs as either problems from Stage 1 or features noticed only at Stage 2. • Failure to establish connections with problems at Stage 1. • Incorporations coded as 'all or nothing' categories. • Lack of a CG to discard task

				repetition effects.
Hanaoka (2007)	• 2 models (picture prompt) • Note-taking	Individual: 37 Japanese EFL university students at higher and lower levels of proficiency	• The students noticed mostly lexical features as they autonomously identified their respective problems, found solutions through models, and incorporated them in their revised texts. • More proficient learners noticed significantly more features than less proficient learners during the comparison stage. • Among the FNs in the model, those that were related to PFNs were incorporated at a higher extent and were also retained longer than unrelated features.	• Note-taking falls short (some incorporations were not reported). • Learners' main concern with vocabulary can be due to note-taking, which oriented them to report lexical items more frequently than other features. • Noticing at Stage 2 is restricted exclusively to problems that are solvable from the model. Unsolvable problems are excluded from the Stage 2 analysis. • Failure to account for unreported noticing from the model that led to incorporations. When this occurred it was taken as evidence of incomplete note-taking. • Analysis is restricted to lexis only. Focus is on quantifying the relationship between noticing and incorporations. • Lack of a CG to discard task repetition effects.
Abe (2008)	• Model (essay) • Think aloud and interview	Individual: 14 Japanese ESL university students at higher and lower levels of proficiency	• The model essays led the L2 learners to notice various aspects of language. • The learners' proficiency levels and the differences of the writing tasks were significant factors in explaining	• The sample size is too small for results to be extrapolated. • Because the two model essays adopted in the study may not be representative of all essays, the

			the difference in quality and quantity of their noticing. • Model essays not only functioned in a similar way to the texts reformulated by native English teachers, but also had unique advantages as feedback tools.	generalization of the finding pertaining to task differences is also problematic. • Lack of a CG to discard task repetition effects.
Yang & Zhang (2010)	• Models and reformulations (picture prompt) • Pair talk and marking	Collaborative: 10 Chinese EFL university students	• The problems experienced at Stage 1 promoted the students' awareness of holes in their linguistic knowledge at Stage 2 with both forms of feedback. • The learners could notice most differences between their original text and the reformulated one at Stage 2. • Reformulations were useful for noticing grammar and lexis, and models for lexis. • Reformulations, models and collaborative work contributed to the improvement of the learners' revised texts.	• Small sample size. • Collaborative dialogue only recorded at stages 2 and 3, thus failing to capture problem-solving processes during Stage 1. • There is no way to see what the • contribution of reformulation and models is in the final individual product. • Lack of a CG to discard task repetition effects.
Martínez Esteban & Roca de Larios (2010)	• 2 models (picture prompt) • Note-taking	Individual vs. collaborative: 17 15-year-old Spanish EFL students	• The students noticed mainly lexical problems at Stage 1 but could only find a few solutions in the models. • The comparison stage allowed them (especially those who wrote in pairs) to notice a large quantity of features related to the content of the pictures and the linguistic means used to express that content.	• Small sample size. • Failure to include solved problems at Stage 1. • Failure to account for unreported noticing from the model that led to incorporations. • The feedback was not evenly useful for all the participants, especially for individual writers. Invidual

			• The number of incorporations into subsequent revisions was found to be high.	differences could be considered in the future to explain these differences. • Lack of a CG to discard task repetition effects. • Lack of justification for the use of two models.
Hanaoka & Izumi (2012)	• Models and reformulations (picture prompt) • Note-taking	Individual: 38 Japanese EFL university students of intermediate proficiency	• Most of the problems reported were lexical. • The students found solutions to their problems and incorporated them in their revised texts regardless of the type of feedback. • The two types of feedback texts played different roles: the model text offered solutions to both overt and covert problems roughly equally, while reformulations provided solutions mostly to overt problems.	• Small sample size • Failure to account for unreported noticing at Stage 2 although this phenomenon is acknowledged as a methodological limitation of note-taking. • Failure to account for the incorporation (or not) of unreported features. • Incorporations coded as 'all or nothing' categories. • Lack of a CG to discard task repetition effects.
Coyle & Roca de Larios (2014)	• 2 models and EC (picture prompt) • Note-taking	Collaborative: 46 10-12-year-old Spanish EFL learners	• Noticing at Stage 1 was mainly lexically driven. • Learners in the EC condition reported more noticing of formal aspects at Stage 2, which later emerged in their revisions. Models directed attention toward lexis and language chunks. • Learners' revisions were mostly	• Small sample size. • Failure to account for unsolved and unreported problems at Stage 1. • Failure to provide an even distribution of pairs of different proficiency levels in each of the feedback conditions.

			lexical, although EC encouraged more grammatical revisions than models. • Gains in the linguistic acceptability and comprehensibility of their revised texts showed an advantage for EC over models. • The children in the model group experienced difficulties in understanding and internalizing the language.	• Incorporations coded as 'all or nothing' categories. • Lack of a CG to discard task repetition effects. • Model texts should have been more carefully adjusted to learners' needs to encourage understanding.
Cánovas Guirao, Roca de Larios & Coyle (2015)	• Models (picture prompt) • Pair talk and note-taking	Collaborative: 20 10-11-year-old Spanish EFL learners	• At Stage 1, the children focused mostly on lexis. • Models were useful for attracting children's attention to lexis and chunks of language. • Learners in both groups made a similar number of overall changes in their stories from Stage 1 to Stage 3, with practically the same number of acceptable changes in both cases, most of which were lexical • Proficiency levels were found to influence noticing and uptake from the feedback.	• Small sample size. • Failure to account for unreported problems at Stage 1. • Individual differences were not taken into account. • Standardized measures of proficiency could also be used to avoid dissimilar levels between groups of learners.
García Mayo & Loidi Labandíbar (2017)	• 2 models (picture prompt) • Note-taking	Individual: 60 13-16-year-old Basque-Spanish EFL learners of two different proficiency levels,	• Learners noticed and incorporated mainly lexical problems, although they also attended to content features. • More proficient learners and guided learners noticed more features.	• Failure to account for the incorporation of unreported features, although this phenomenon is acknowledged as a methodological limitation of note-

		divided in turn into three groups (CG, guided and unguided noticing)	• Overall, learners had a negative attitude toward writing and modeling, but those with more positive beliefs incorporated more items in subsequent revisions.	taking. • Failure to focus on features incorporated but not related to Stage 1.
Coyle, Cánovas Guirao & Roca de Larios (2018)	• Models (picture prompt) • Pair talk and note-taking • Longitudinal study	Collaborative: 16 10-11 Spanish EFL learners, some of which received prolonged instruction, and some did not	• Children followed sequential routes across writing and feedback processing tasks that involve a wider range of interrelated problem-solving behaviours, noticing and uptake possibilities. • The children's use of trajectories was found to impact on their developing L2 knowledge. • Children who received instruction with models followed more beneficial trajectories than their peers.	• Failure to provide statistical analysis of the data given the small number of participants • Improvement in the teaching group could be attributed to task-repetition effects. • The fact that the researcher was also the class teacher may have influenced the quality of the research project. The true picture may not be provided because of personal evaluation by the teacher.
Montealegre Ramón (2019)	• Models (picture prompt) • Note-taking	Individual: 30 13-14-year-old Spanish EFL learners, divided into two different types of noticing groups	• There were no differences between the two groups within stages. • All the students increased the number of features reported across stages regardless of the feedback condition. • The model text provided the learners with alternative features related to lexis, form, and ideas. • The rewriting stage resulted in a similar number of changes across categories in both groups.	• Small sample size. • Note-taking is not always successful in capturing the thought processes. • Failure to account for individual differences.
Lázaro-	• 2 models	Individual vs.	• All students noticed and incorporated	• Small sample size.

Ibarrola (2021)	(picture prompt) • Note-taking	collaborative: 33 11-12-year-old Spanish EFL learners	few features, and most of them were lexical. • The number of incorrect incorporations was high in both groups. • There was no improvement between the initial and final drafts, but when rated holistically, the students' second draft, was significantly better than the first one. • Individual writers noticed a higher number of features, but pairs were more able to incorporate features correctly.	• Incorporations coded as 'all or nothing' categories. • Lack of justification for the use of two models.
Kang (2020)	• 2 models (writing prompt) • Note-taking	Individual: 40 16-17-year-old South Korean EFL learners	• At stage 1, students mainly noticed lexical problems, which led to a lexically-oriented search for solutions in the model. • Some of these words were incorporated into their rewriting. • The students in the TG received higher writing scores on the posttest compared to those in the self-correction group due to the incorporations. • Model texts were found to be particularly effective in improving vocabulary and content of L2 learners' writing.	• Note-taking falls short. • Failure to account for individual differences. • Lack of justification for the use of two models.
Luquin &	• Models (picture	Collaborative: 12 11-	• Unlike previous research, the children	• Small sample size.

García Mayo (2020)	prompt) • Pair talk	12-year-old Spanish EFL learners	noticed mostly grammatical problems at the writing stage. The authors attribute these results to a grammar-oriented teaching approach. • At stage 2 both groups focused most of their attention on content and lexical features. • The learners in the TG noticed in the model text and incorporated in subsequent writing significantly more lexical features than the CG.	• Failure to account for individual differences. • Focus only on the type and quantity of categories noticed and incorporated.
Coyle & Roca de Larios (2020)	• 2 models (picture prompt) • Pair talk • Note-taking at stage 2	Collaborative: 16 Spanish EFL (10-11-year-olds) and CLIL (9-10-year-olds) learners from two different schools	• Both the CLIL and EFL pairs focused primarily on lexis. • Four strategies were observed in the data, but differences were found in the extent and nature of their use by pairs from each context, and in their impact on the children's uptake. • While noticing surface-level differences between the model and their own draft was the most used strategy by all pairs, only the CLIL pairs paid attention to new and alternative features in the feedback. • A great amount of the noticing from the models which was not reported in the pair talk appeared in their revised texts.	• Small sample size • Written notes were only asked at Stage 2, thus failing to complement the children's verbalizations at stages 1 and 3 • Lack of a CG to tease out task-repetition effects • Failure to administrate children a standardized proficiency test. Although the authors consider that both groups are comparable, the variables age and proficiency level may have offset the advantages of one group over the other
Luquin & García	• Models (picture prompt)	Collaborative: 38 11-12-year-old Spanish	• The TG noticed a considerably higher number of features than the CG.	• Failure to include discussion of problems arisen at Stage 1

Mayo (2021)	• Pair talk	EFL learners	• The TG noticed significantly more lexical and content-related features than any other category and the difference was also significant with respect to the CG's noticing of the same features, while the self-correction group focused mainly on form and mechanics. • At Stage 3 and 4, the dyads in the TG incorporated significantly more features than their counterparts, most of which were discursive, mechanical and formal.	• Small sample size • The study focused on what YLs noticed and incorporated, but not on how accurately they worked out their problems throughout the four stages • Failure to provide more information on the cognitive processes that intervene in the 'unblocking' of new knowledge • Possible procedural task repetition effects
Villarreal & Lázaro-Ibarrola (submitted)	• Models (picture prompt)	Collaborative: 32 9-10-year-old Spanish CLIL learners	• Regarding text quality, results showed a significant increase in complexity in the MG. • Motivation ratings in both groups were generally positive and showed a significant increase post-task. • Working with their peers was the most selected reason for their high scores. • Motivation levels, on the other hand, differed considerably at the post-test: while the CG maintained them, the MG evidenced a significant decline.	• Small sample size • Failure to further develop the reasons for the negative disposition of the children in the MG. This would also allow future investigations to test and refine them

3.4. Conclusion

As can be seen, the research summarized above has provided empirical evidence for the benefits of using models as an alternative form of feedback with students of different ages and proficiency levels. The findings lay bare that students are able to and actually do notice gaps between their written output and the feedback, and that this noticing seems to have a beneficial impact on the use of the TL, at least in the short-term. The results of the studies on models conducted with adults do not differ much from those carried out with children. Overall, models seem to attract most of learners' attention to content, lexical features, alternative expressions and ideas, and most of the incorporations are also lexical in nature. This leads us to think that modeling is not a good option for the noticing of formal aspects of the language. For one thing, we cannot forget that models are usually operationalized through the use of a picture-based descriptive task, which inherently pushes students to describe actions. While performing the task, children struggle to find the right nouns, verbs, adjectives, and adverbs to describe the picture prompts, thus directing most of their attention to lexis and content (Hanaoka & Izumi, 2012; Kang, 2020). In addition, this focus on lexis might also be attributed to the limited processing capacity of the human brain which cannot process content and meaning simultaneously and therefore favors meaning over form, especially at low levels of proficiency (VanPatten, 1990, 2004; see section 4.2). As for task completion, the optimistic findings attained as a result of the combination of collaborative writing and models are consistent with the general benefits evidenced for collaborative writing and models independently and motivate more research to see if the gains are certainly greater in the collaborative modality. Finally, we have seen that many studies code the students' incorporations as either accurate or not. This methodological decision completely misses out all the partial improvements and minor gains that YLs can obtain from WCF and which have been an important contribution of previous research (e.g., Cánovas Guirao et al., 2015; Coyle et al., 2018; Yang & Zhang, 2010). The use of 'all or nothing' criteria on the students' written output (and even more so with children) does a disservice to the learners' use of language and fails to consider changes in the learners' texts including partial corrections, substitutions, new incorporations, etc. Much more nuanced approaches are needed to provide true insights into the impact of models on written output.

Consequently, further research is needed to know more about the merits of this alternative technique and to ascertain that teachers will be able to implement them in the language classroom as a 'writing-to-learn' resource. As such, this dissertation contributes to our increasing knowledge about the language learning potential of model texts as an alternative form of WCF which appears to play a pivotal role in promoting noticing and metalinguistic awareness, and in engaging students in deeper processing than with more traditional CF techniques. What remains to be discovered in more detail is if this potential can foster advantages over a longer period of time and with linguistic features other than lexis. To round off this research, we consider that individual variables also have a say in the processing of any learning, and language learning does not deserve any less. Issues such as developmental readiness or motivation are crucial in the field of child SLA, and as such, they are certainly worth due consideration. Therefore, the next and last chapter of the literature background will be devoted to the individual learner.

CHAPTER 4: THE ROLE OF INDIVIDUAL DIFFERENCES IN L2 WRITING: ONE SIZE FITS ALL?

There is a large number of studies that describe the effects of individual learner differences regarding L2 speaking (e.g., Dörnyei & Kormos, 2000; Kormos & Trebits, 2012) and reading skills (e.g., Grabe, 2009). Little is known, however, about their impact on L2 writing, on the quality of the written production, and on students' response to WCF (Bitchener, 2012; Kormos, 2012). Individual learner-internal factors, together with contextual factors, mediate the way learners engage in feedback processing thus causing variations in L2 writing practices. These internal factors encompass cognitive factors such as working memory, developmental readiness or their level of L2 proficiency among others, and motivational/affective factors such as goals, interests, attitudes or beliefs (Bitchener & Storch, 2016). The interplay between these variables is likely to determine the result of feedback processing. As Kormos (2012, p. 390-391) observes:

> Writing is not only a complex task but also a time-consuming activity that requires concentration and determination. Producing 100 words orally might take about a minute in an L2, whereas writing a composition of 100 words might take 30 minutes. Furthermore, one often has no choice but to respond orally in communicative situations; composing a text, however, might frequently be an optional activity. All these characteristics of writing suggest that a learner's motivational profile and self-regulation may play a very important role in determining whether students will engage in writing activities, what kind of writing tasks they will undertake, with what level of effort and attention they will approach the various phases of the writing process, and how they exploit the learning potential of writing tasks.

All these factors should receive special attention when it comes to YLs, since many cognitive, social and physical changes take place in the developing child (Nicholas & Lightbown, 2008). Research that has specifically addressed the degree at which these factors might influence the engagement of students with WCF will be discussed in the next four sections. In each section we will first describe the factor and contemplate how they may facilitate or hinder the child' cognitive processing of WCF. Subsequently, we

will provide an overview of some of the research conducted on the corresponding individual difference.

4.1. Developmental readiness

As discussed in chapter 2 (see section 2.1), children may not be developmentally ready to completely take advantage of all the benefits that WCF offers, and therefore their processing ability is limited. Pienemann's (1989) Teachability Hypothesis claims that language can only be successfully acquired when learners are at the right stage of IL development to acquire them, because L2 acquisition is supposed to follow a fixed sequence that is immune to external mediation such as feedback. Likewise, Truscott (1996) argued that for CF to have any potential impact, it should be aligned to the learner's current level of L2 development to avoid providing students with formal structures that they would not yet be ready to acquire.

On the other hand, it has also been put forward that what is already known regarding developmental stages of acquisition is too limited to be applicable to educational settings (e.g., DeKeyser, 1998; R. Ellis, 1997; Truscott, 1996). This led Truscott (1996) to conclude that even the provision of aligned CF cannot be expected to be of help to language acquisition. Contrary to Pienemann's predictions, studies on the developmental stages for English questions (Mackey, 1999; Spada & Lightbown, 1999) showed that children who were ready to advance tended not to do so, while some 'unready' students progressed more rapidly through the stages. In particular, Spada and Lightbown (1999) tested the developmental progress of 150 11-12-year-old French-speaking ESL students in the use of English questions before and after a series of focused activities on question formation. The results revealed that form-focused instruction and oral CF aided the learners in moving forward sequentially in their use of English questions. In a like manner, Mackey (1999) and Mackey and Silver (2005) observed that some adult and child ESL students that were not supposed to be ready to acquire the targeted forms were able to advance faster to the stage immediately following their current developmental level, but they were also unable to avoid any of the stages. In view of these findings, explicit developmentally appropriate instruction

and potentially also WCF would seem to accelerate language learning without altering it considerably.

Relatedly, a series of morpheme order studies (Dulay & Burt, 1974; Krashen, 1977) conducted in the 70's in the area of SLA set the basis for the Natural Order Hypothesis (Krashen, 1985), which claims that children acquire grammatical structures in a pre-determined, 'natural' order. Such studies revealed that the order of acquisition of certain morphemes including the present progressive (*–ing*), past irregular, plural (*–s*), possessive (*–'s*), copula, articles (*a*, *the*) or past regular (*–ed*), to name but a few, was similar for younger and older learners in naturalistic settings irrespective of their L1 background. However, the Natural Order Hypothesis was investigated empirically in instructional settings obtaining conflicting results. For example, Pica (1983) worked with adult Spanish L1 speakers in the United States who were grouped into three different acquisition settings: naturalistic, instructed and mixed (a combination of classroom instruction and natural exposure to the TL). She observed that the morpheme rank orders were similar for the three groups, providing support for the claims made by previous researchers to this effect. Likewise, the morpheme order was also found to be maintained in the case of children and adolescents enrolled in formal EFL instruction in Japan (Makino 1979, as cited in Pica, 1983). On the other hand, in Sajavaara's (1981, as cited in Pica, 1983) study, a disturbed natural order was found among instructed Finnish L1 learners. For her part, Lightbown (1983) concluded that the differences found in the accuracy order of Francophone EFL children in Quebec were caused by the distorted version of English the children were exposed to, and the absence of communicative language practice, which resulted in little improvement over time in their use of the TL. Nevertheless, upon reviewing the abovementioned morpheme studies to evaluate the view that there is a universal order of acquisition, recent research has demonstrated the strong influence of the L1 on the acquisition of grammatical morphemes. For instance, Murakami and Alexopoulu (2016) analyzed the L2 acquisition order of six English grammatical morphemes by students from five different proficiency levels. The authors found that the order of acquisition is not insensitive to L1 knowledge. What is more, they reported that those morphemes containing semantic concepts (interpretable) are less impervious to L1 influence than those encoding language-independent concepts (uninterpretable). For their part, Pei-sui Luk and Shirai (2009) reviewed the literature to look into whether or not native speakers of Japanese, Korean, Chinese, and Spanish

followed the natural order in their acquisition of English morphemes. Findings revealed that native speakers of Japanese, Korean, and Chinese showed a later acquisition of plural –s and articles, and earlier acquisition of possessive *'s* than predicted by the natural order.

The analysis of morphological order has also been researched in a series of studies with instructed Spanish/Basque EFL learners (García Mayo & Lázaro-Ibarrola, 2005; García Mayo, Lázaro-Ibarrola & Liceras, 2005; Lázaro-Ibarrola, 2002, as cited in Lázaro-Ibarrola & García Mayo, 2012). The authors observed that the morphological development of these learners accelerated at around age 12, possibly as a result of the acquisition of the pronominal system which is completely achieved at around age 15. Later, Lázaro-Ibarrola and García Mayo (2012) examined the morphosyntactic development of a group of CLIL adolescents while they were narrating a story at two different points in time (age 13 and 15) and observed significant development over time. What is more, they found that their learners seemed to be more advanced when compared to similar data from non-CLIL students. In light of these results the authors deduce that because of CLIL instruction or other factors, the learners were able to surpass the ceiling effect reached by their non-CLIL classmates.

Considering the research conducted so far on developmental readiness, Ortega (2009b, p. 138) concludes the following:

> True, for some developmental areas, such as sequences for word order […] and tense and aspect morphology […], learners appear psycholinguistically unable to skip stages. But for other areas of the grammar, instruction above the cutting edge of a given interlanguage may accelerate development.

Developing this idea further, Izumi (2013) contemplates the possibility that the metalinguistic awareness accomplished through explicit instruction might help students improve their noticing which, in turn, seems to play an important role in enabling students to move along stages. These perspectives underpin the general opinion that whereas instruction and (W)CF may not be powerful enough to alter or modify the order of acquisition, they appear to have a positive effect on the speed at which L2 grammar is acquired.

4.2. Working memory

The learner's working memory, also known as short-term memory, is the place where new information is held temporarily until it is either lost or placed into long-term memory. It is the part of memory where such crucial processes as attention, noticing, hypothesizing, restructuring and practice take place (Bitchener & Storch, 2016). As opposed to long-term memory, the working memory has a limited capacity and is thus restricted by the amount of information each individual is able to process at any one time (Bitchener & Storch, 2016). Consequently, the degree at which students notice, understand and incorporate new L2 features from WCF into their IL system will be determined for the most part by their ability to process information. The limited capacity model (e.g., Skehan, 1998) argues that the learner's working memory is of limited capacity and while some may experience attentional constraints during the processing of feedback, some others may have larger working memory capacities and are therefore better endowed to pay attention to and process input. According to this model then, tasks that are excessively demanding absorb learners' attention and memory resources in such a way that students deviate their attention to message content rather than to linguistic form. Pertaining our field of inquiry, Michel et al. (2019) observed that writing and feedback analysis entail complex interactive and recursive cognitive processes. Similarly, Kormos (2012) explained that individual differences in working memory could even determine how L2 learners handle and store information during written production. Although the written mode is not constrained by time as speech is, it still entails the concurrent retrieval, storage and handling of language and ideas to be worded. In developing writers, such as YLs, even the automatic processes of writing letters demand certain attention and mental effort. To this we should add that during editing, writers need to read what has been written and simultaneously pay attention to coherence, cohesion, accuracy, and appropriateness (Kormos, 2012).

This would seem to be particularly true for less knowledgeable learners to whom the high working memory load as well as their limited L2 knowledge would place heavy demands when it comes to (i) allocating their attention to more than one aspect of language at the same time, (ii) noticing holes and gaps in their drafts and filling them with the feedback, (iii) encoding linguistic form and structure, and (iv) based on the new information, formulating and testing hypothesis about correct language use

(Cánovas Guirao, 2017). Therefore, these learners would need to process new input in a more consciously controlled manner (Bitchener & Storch, 2016).

In this sense, individual differences in working memory are notably significant for children whose attentional regulation mechanisms are still developing (Michel et al., 2019). According to Izumi (2003), it is possible that YLs might find difficulties during feedback processing as their limited L2 knowledge may push them to look at semantic and contextual clues, thus holding back attention to form. Consequently, children are likely to only process input in a shallow way, as deep processing involves greater awareness, more attention and more cognitive effort to establish significant relationships between the TL and their IL (Leow, 2015). Therefore, highly demanding tasks do a disservice to children with lower levels of working memory operating in high-stakes and classroom testing contexts (Michel et al., 2019).

In his 'primacy of meaning' principle, VanPatten (2004) maintains that learners process input for meaning before they process it for form. In summary, the basic tenets of this hypothesis are that (i) learners process content before anything else, (ii) morphology and syntax are likely to be disregarded to the detriment of lexis, (iii) learners rely on lexical words to encode meaning before grammatical forms that indicate the same semantic information (e.g., *tomorrow* before *will*) and (iv) semantic features are processed before less communicative forms (e.g., *he* before third person *–s*). What this implies in studies with YLs is that children 'may not attend to and notice a particular structure if they are processing the L2 input primarily for meaning and this attention to meaning exhausts most of their resources' (Uggen, 2012, p. 511).

There is a scarcity of research on the role of working memory in L2 students' writing processes and attainment, and far less in children's. Some exceptions are the study of Ndlovu and Geva (2008), who worked with children with specific learning difficulties in a Canadian context. They compared the writing skills of L1 and L2 children, who were assessed as being reading disabled or non-reading disabled. They observed that, irrespective of their language background, the children with specific learning difficulties struggled with lower-level writing processes including spelling, punctuation, and the monitoring of syntax. These learners also exhibited problems with sentence structure, the generation and coordination of vocabulary, and with the composition of stories.

Another more recent study in this area is that of Michel et al. (2019), who looked into the role of working memory in the writing performance of 94 EFL Hungarian children aged between 11 and 12 years old. In general, the findings revealed that children with high working memory functions showed more consistent performance across tasks than those students with low working memory functions.

4.3. L2 proficiency level

The successful processing of WCF is also contingent on the learner's level of proficiency in the L2. Actually, studies on input and output processing tasks have attested that students' performance is mediated by their knowledge of the L2. In this way, while high proficiency learners have been found to perform written tasks and process input with high levels of attention, low-proficiency learners seem to benefit less from the input. This is exemplified in the work undertaken by Qi and Lapkin (2001), who explored the language learning potential of reformulations with two Chinese ESL adult learners at different proficiency levels. They observed that the more advanced writer noticed a higher number of gaps from the reformulated text, and that the quality of this noticing was different between the two of them. That is, the learner with a higher proficiency level was better able to verbalize the reasons for accepting a reformulated item than their less advanced counterpart. In view of these results, the authors suggest that low proficiency learners may have more difficulty understanding the nature of the gaps they notice from feedback and so benefit less from the reformulation. Two Japanese EFL adult learners at different levels of proficiency participated in Hanaoka's (2006b) study. She also found that the more proficient learner noticed more features from a reformulation and a model than the less proficient learner, although the number of incorporations from the feedback texts was about the same. Nevertheless, it is important to bear in mind that the findings reported by Qi and Lapkin (2001) and Hanaoka (2006b) are based on the data of only two learners, so these results need to be interpreted with caution.

Research conducted with younger learners has yielded similar results. On the subject of models and/or reformulations, Coyle and Roca de Larios (2014) found that the less

proficient children noticed more problems with their drafts at the comparison stage than the higher-level children, which could be interpreted as an aftereffect of going through a higher amount of difficulties during text production. Cánovas Guirao et al. (2015) reported that the more advanced learners noticed and incorporated more features from the model than the less proficient ones. Regarding dictogloss tasks, this time with older YLs (12-13-year-olds), Lapkin, Swain and Smith (2002) also observed richer collaborative dialogues among the stronger pairs while they were engaged in jigsaw and dictogloss tasks. In addition, the higher proficiency dyads provided a more detailed noticing than their weaker classmates and created more contexts for the pronominal verbs, which were the focus of the tasks.

In relation to form and following VanPatten's (2004) 'primacy of meaning' principle, Leeser (2004) concludes that more proficient learners would find it easier to process grammatical form than lower-level students on the grounds that the former would not have to struggle as much with processing meaning. Likewise, as seen above, the literature on 'developmental readiness' puts forward that students would be ready to acquire a certain grammatical form only when they could produce less complex structures. Again, Leeser (2004) implies that more advanced learners should be developmentally ahead of less proficient learners to notice and produce specific structures during communicative tasks. Weaker learners, on the other hand, may experience difficulties focusing on form while they are trying to understand meaning, and if they manage to do so, they will rely on those forms that carry most of the meaning. In relation to students' languaging, Swain (1998, p. 73) observes that 'learners talk about what they need to talk about, that is, those aspects of language about which they are not sure. And that, in turn, will depend on their own current, internalized state of knowledge about language and its use'. Based on this idea, proficiency in the L2 would also seem to exert an influence on the types of LREs (Swain, 1995). By way of example, the interaction between proficiency and the types of LREs produced during a collaborative task has been investigated by Leeser (2004). He analyzed how different proficiency grouping (proficiency-matched and mixed proficiency dyads) influenced the amount, type and resolution of the LREs produced during a dictogloss task performed by 42 L2 Spanish learners (L1 English). The author reported that the more proficient learners produced more LREs, that pairs with high proficiency students generated more grammar-related LREs than lexical LREs, and correctly resolved a higher number of

LREs than the other two pairs. These findings were confirmed by Basterrechea and Leeser (2019) in a study conducted with 24 EFL adolescent learners on the production of the 3rd person singular *–s* in a CLIL context: high-proficiency learners generated more LREs, attended to form more frequently and resolved a higher number of LREs than the lower-level students. Nevertheless, it is important to mention that the fact that stronger learners are able to capture a higher number of (grammar-related) LREs due to their cognitive maturity does not imply that YLs, even at a beginner level, cannot focus on formal issues, as was the case in studies such us Calzada and García Mayo (2020a, 2020b) or Luquin and García Mayo (2020a, 2020b).

4.4. Motivation

In the previous sections, we addressed the effect of cognitive factors on the processing of writing and WCF, but there also exist individual learner-internal factors that have to do with the affective engagement with WCF, such as learner's motivation. According to Bitchener and Storch (2016), motivation would appear to be vital for the consolidation of the student's new L2 knowledge, so that it can be retrieved automatically from their long-term memory in the course of time. Their personal experiences in the classroom with writing and WCF are likely to mold their motivation and beliefs concerning the value of these activities (Bitchener & Ferris 2012, Manchón, 2011). In addition, in contrast with cognitive factors, which are insensitive to social and instructional agents, motivation is usually affected by the complex interplay of the social, cultural and educational contexts (Kormos, 2012). Therefore, gaining some insights into the motivational justification for the differences in learners' attitudes toward and performance in WCF tasks is paramount, since 'it is hard to imagine learners with little or no motivation for writing in a second language to care about learning from WCF' (Waller & Papi, 2017, p. 55).

The importance of motivation in the efficacy of WCF has been attested by many researchers in the field (e.g., Dörnyei & Kormos, 2000; Ferris, 2010; Goldstein, 2005; Hyland, 1998, 2011; Kormos, 2012). Dörnyei and Kormos (2000), for example, put forward that students are more inclined to communicate if they have a positive attitude

toward the task. Similarly, Goldstein (2005) claimed that low attention to WCF could be ascribed to students' lack of motivation. In Hyland's (2011) analysis of university learners' attitudes and motivation to learn from the feedback, she proves that their willingness to engage with WCF is largely dependent on their learning goals. Kormos (2012, p. 399) rightly summarizes the effect of motivation on writing and feedback processing:

> Noticing gaps in one's knowledge and engaging in problem-solving behaviours that can potentially promote acquisition processes require increased motivational effort, intrinsic interest in language learning, and positive self-efficacy beliefs. If students lack strong goal orientation and interest and do not believe in their ability to successfully acquire the L2, then they are only likely to complete the writing task itself and may not engage in further cognitive processing or collaborative effort to learn from the task. Similarly, motivational intensity also affects learners' attention paid to feedback and their further involvement in creating text revisions.

Regarding feedback forms, two studies have examined learners' affective engagement with some types of feedback such as reformulations (Storch & Wigglesworth, 2010) and EC (Hyland, 2011). Nevertheless, to our knowledge, adult and adolescent learners' attitudes toward models have only been explored in four studies. In Hanaoka (2007), the undergraduate participants were asked to indicate how eager they were to read the models on a scale of 1 to 5. The mean score was 4.3, revealing that the participants were strongly motivated to receive the models in the following stage of the task. In a similar vein, Yang and Zhang' (2010) university participants were asked to write down their comments and were also interviewed to elicit their attitudes toward this method. What the authors found was that the students valued the nativeness of the models as well as the fact that this form of feedback enabled them to engage in deeper reflection beyond their own texts. On the contrary, some students missed further assistance regarding the selection of vocabulary, choice and ordering of syntactic structures, markers of cohesion, and discourse functions found in the model texts.

These researchers, however, did not explore how the participants' motivation could influence their performance in subsequent revisions. García Mayo and Loidi Labandibar (2017) addressed this gap by administering a background questionnaire on the use and

perceived effectiveness of models to their adolescent participants and comparing the drafts of the motivated and less motivated students. The learners' answers revealed that although they found the models quite useful, they did not really enjoy them. In addition, the authors found that highly motivated students tended to incorporate more features in subsequent revisions, while a low motivation behavior gave way to a poorer performance. In view of these results, García Mayo and Loidi Labandibar (2017) conclude that their participants' negative attitudes toward modeling and writing could be attributed to the deep emotional changes that their adolescent learners were undergoing, in comparison with the high levels of motivation found in Hanaoka's (2007) and Yang and Zhang's (2010) undergraduates. Although we do not know about whether or not her participants enjoyed studying the models, Kang's (2020) adolescent learners perceived modeling as somewhat helpful in improving their writing. Additionally, they stated that models allowed them to learn new vocabulary and expressions, become familiar with the organizational structure of argumentative essays, identify problems in their own writing, or improve their understanding of text coherence. However, some students opined that the models were not enough to correct their errors, and that they would have preferred to receive explicit corrections.

It is commonsense to think that children and motivation should go hand in hand, especially in classroom contexts where YLs have to deal with arduous and demanding tasks on a daily basis (Calzada & García Mayo, 2020b; Shak, 2006). Children should find learning enjoyable, in such a way that they can develop a sustained level of motivation necessary for long-term achievement, as Bitchener and Storch (2016) explain. Lack of motivation, on the other hand, is an important barrier to academic success, productivity, and wellbeing over time (Legault, Green-Demers & Pelletier, 2006). An example of research into children's motivation is the work done by Lo and Hyland (2007) in Hong Kong. They looked at the attitudes of 40 10-year-old ESL writers on a new writing program designed to foster their motivation and engagement. This program consisted of tasks which were relevant to students in the sense that the topics had to do with their life experiences and the children were also provided with a real and wider audience and a real purpose for writing. In looking at the interviews with focus groups as well as the children's log entries and short questionnaires, the researchers found that the students showed high levels of motivation and engagement. Writing for genuine audiences on topics of interest and relevance to them increased

their involvement with the task and also led to improved performance and longer texts, especially for less able students, although accuracy and organization scores fell. For the authors, these findings suggest that the enthusiasm shown by the children compensates for any short-term declines in accuracy which, as they suggest, could be solved with more scaffolding.

A more recent study is that by Azkarai and Kopinska (2020). They analyzed the relationship between motivation, level of engagement in LREs and patterns of interaction in a group of Spanish EFL students aged between 11 and 12 years old who worked collaboratively on a dictogloss task. To serve that objective, the participants were asked to complete a motivation thermometer (adapted from Al Khalil (2016)) before and after completing the task. This thermometer asked them to rate how motivated they were to do the activity on a 10-point scale and to underline the reasons why. What Azkarai and Kopinska (2020) observed is that the initial positive attitude toward the task increased after finishing it, but this increase was only significant in the case of the cooperative pairs. Also, no direct relationship was found between the level of engagement in LREs and task motivation.

Studies about children's attitudes to WCF are much less abundant. An exception is Fazio's (2001) analysis of the impact of different feedback techniques (corrections, commentaries and a combination of the two) on the journal writing accuracy of 112 10-11-year-old learners of French in Canada. The author reported that none of the feedback conditions seemed to promote changes in the students' accuracy. Fazio (2001) attributed this result to the learners' passive attitudes toward the provision of feedback by someone other than their familiar classroom teachers, but it was also observed that learners did not attend regularly to the corrections provided. The author related these issues to emotions, aspirations and secrets found in the children's journal entries which overshadowed attention to corrections. First and foremost, what this study stresses is the idea that noticing and attention, which are necessary conditions for CF to be successful, are usually contingent on contextual and attitudinal variables.

To our knowledge, child learners' level of motivation when dealing with models has only been explored by Lázaro-Ibarrola and Villarreal (2021) and Villarreal and Lázaro-Ibarrola (submitted). Lázaro-Ibarrola and Villarreal (2021) analyzed the pre- and post-

task motivation scores and the motives of 12 dyads of CLIL YLs (aged 9-10) divided into a model group (MG) and a task repetition group (TRG). The children in both groups wrote collaboratively during three sessions. In the first session, all the pairs wrote a composition; in the second session, the MG was provided with model texts while the TRG re-wrote their drafts, and in the last session, all the learners wrote the same text again. Before and after each task, the students completed an individual motivation thermometer by choosing a (1-10) score and a reason for it from a list selected. The results revealed that the overall children's motivation was high. However, while the TRG maintained their motivation throughout the tasks, the positive disposition of the learners in the MG decreased when they had to rewrite their initial draft trying to incorporate the features noticed from the model. The authors suggest that the use of models had a demotivating influence on the children, as opposed to the mere repetition of a draft. Concerning motives, working with their peers was the main reason the learners provided to justify their high motivation.

The latter study by Villarreal and Lázaro-Ibarrola (submitted) looked at the effects of models on the collaborative writing of 13 pairs of CLIL children aged between 9 and 10 years old and addressed both draft quality measures and learners' motivation. The children were divided into a CG (who received no feedback) and a MG (who received feedback in the form of models) and engaged in a three-stage task. Before and after each task, all dyads filled out an individual motivation thermometer in which they had to rate their motivation and select their reasons for their ratings in the thermometers. The pre- and post-motivation scores were compared, and motives were analyzed qualitatively. Regarding motivation, the findings revealed that the scores in both groups were mostly positive and showed a significant increase post-task, with collaboration being the most selected justification for their high rating. Motivation levels, nevertheless, differed considerably at the post-test: while the CG maintained them, the MG exhibited a significant drop.

As can be seen, motivation may, in general, operate at an emotional level and be determinant for language learning processes. In particular, motivation might have an influence over students' disposition to focus on the accuracy of their writing, and therefore to WCF, and to engage in cognitive processes (Bitchener & Storch, 2016). What is more, if students are provided with a specific feedback form and believe that

the feedback has no value or interest for them, they may decide to disregard it by not attending to it and not cognitively processing it. That could be the case, for example, if prior experience with that type of feedback was not successful, in which case these types of experiences could be one of the causes for their demotivation (Bitchener & Storch, 2016). Although a great amount of research has been carried out on motivational factors affecting written tasks, to our knowledge, there are no studies that investigate both quantitatively and qualitatively the role of motivation in EFL children's collaborative writing and feedback processing.

4.5. Conclusion

In this chapter, we have addressed the possibility that the path toward L2 learning, or more specifically toward WCF processing, could also be facilitated or interrupted by intervening learner internal (cognitive and motivational/affective) factors. Among the cognitive variables, we discussed the role of L2 proficiency, working memory and developmental readiness, which have been found to mediate the way learners engage with feedback processing causing variations in L2 writing practices. Cognitive factors, however, also interplay with affective variables, and they separately and jointly impact on writing processes. For the purpose of this dissertation, we focused our attention on motivation, since it seems to be fundamental for the students' consolidation of their new knowledge (Bitchener & Storch, 2016) and is also sensitive to social, cultural and educational factors (Kormos, 2012).

All these individual differences have been found to affect how learners process feedback, the extent to which they notice gaps and, accordingly, how they take advantage of the learning opportunities provided by writing. Researchers (e.g., Ammar & Spada, 2006; Bitchener & Ferris, 2012; Ferris, 1999) repeatedly point out that 'one size does not fit all' but rather feedback should be carefully adjusted to every learner's attitudes, knowledge, beliefs, problems and desires inasmuch as possible (Murphy & Roca de Larios, 2010). Nevertheless, considerably more work will need to be done to determine how individual variation shape acquisition. For instance, there is a lack of longitudinal studies that analyze the extent to which motivation can be sustained over

time (Bitchener, 2012; Kormos, 2012). Also, information on the role of motivation could be especially practical for designing motivating and engaging writing and assessment activities. Cognitive and affective variables might also have an effect on how students work on writing tasks in collaboration. Consequently, further research could also examine more closely the relationship between individual differences and collaborative writing processes. Certainly, writing is a complex and time-consuming activity which requires concentration, even more so for YLs who are still developing cognitively, physically and socially. Therefore, considering such factors when conducting research with children seems of paramount importance if we want to draw a complete and more accurate picture of what is going on during every child' acquisition of an L2.

SUMMARY OF PART I

The first part of the dissertation has provided theoretical and empirical arguments that support the language learning potential of WCF and all the elements associated with it. The first chapter analyzed the importance of conducting studies on children's L2 acquisition. In particular, we highlighted that much more information is needed into YL as more and more schools are introducing the FL at an ever-earlier age. There is a relatively small body of literature that is concerned with what the developing child actually does when faced with specific tasks that facilitate the learning process. This type of research will be beneficial not only to increase our theoretical knowledge about how children acquire a FL, but it will also have direct implications for educational policies and pedagogical decisions that are adopted for this age group. In consequence, we support the idea that studies on YLs in contexts of FL acquisition in the classroom should be one of the primary research objectives in the coming decades. To date, most research has focused on the analysis of oral interaction among adults and among children while they completed communicative tasks in ESL or EFL contexts. However, research into issues such as how EFL children collaborate in writing, process feedback, focus on formal aspects of the language or how motivation affects their performance is basically non-existent in our context.

In chapter 2 we provided the theoretical rationale supporting the view that WCF can be of help for L2 development. This theoretical perspective rests on both cognitive and sociocultural constructs on learning. In the first section we took a stance on the writing-to-learn strand of research, provided a definition for WCF and summarized the most important criticism CF has received over the years. The second section was devoted to the cognitive theories which integrate leading SLA concepts and hypotheses such as the Interaction Hypothesis, input, output, noticing, Skill Learning Theory or the need for a return to FonF within a communicative context along with the importance of learner initiative (i.e., production of LREs) in collaborative tasks. SCT and collaborative writing were examined in the third section. These constructs also furthered our insights into the language learning potential of WCF by means of the significance adhered to cognitive development as a socially situated activity. In the last section, we brought

together the theories outlined in the chapter in order to explain how the interplay between these constructs may facilitate L2 learning.

Chapter 3 presented a more practical perspective of WCF by zooming in on the nature of the different types of WCF that can be found in the literature. We also addressed the limitations of traditional feedback techniques and explained the reported benefits of modeling as an alternative, and more discursive feedback technique than EC. Finally, we presented a review of the literature on model texts. Models are consensually considered to be valuable for providing learners with a broad range of language input, particularly with new lexis, content, expressions and ideas that they use in subsequent texts. While other forms of feedback only pay attention to superficial aspects of the learner's output and explicitly signals the students' mistake preventing them from any type of cognitive effort, models supply students with input that allows them not only to attend to form, but also to engage in deeper reflection and discussion above and beyond their own text. The results of the studies on models conducted so far with children do not differ much from those carried out with adults and adolescents, but more in-depth studies with YLs are needed to determine whether the benefits associated with models lead only to gains in written accuracy or to language development over a longer period of time.

Finally, in chapter 4 we saw the importance of individual differences in L2 writing. More specifically, we described how learner-internal factors can impact on students' response to WCF. Within these internal factors we examined cognitive variables such as working memory, developmental readiness or their level of L2 proficiency. Cognitive factors, however, also interact with affective variables such as motivation, which seems to be crucial for the students' consolidation of their new knowledge. Together or separately, learner-internal factors along with contextual variables have the power to facilitate or hinder learners' engagement with feedback processing and are therefore likely to determine the result of each learner's performance. What is more, the complex and time-consuming nature of writing insinuates that the student's attitude and cognitive capacities may play a fundamental role in ascertaining whether or not learners will engage in writing tasks, what types of writing activities they will be willing to perform, their level of effort and attention during the writing process, and how they take advantage of the learning potential of these tasks.

Taken together, research on children facing alternative feedback techniques in combination with collaborative writing and which also considers learner-internal factors is necessary, and this dissertation aims to be a small contribution to the overall field of inquiry.

PART II

THE STUDY

CHAPTER 5: METHOD

5.1. Rationale for the study

Throughout this review of theory and research into the language learning potential of collaborative writing and WCF certain questions have been raised that need to be dealt with to fully comprehend the paramount role that collaborative writing and feedback may play in the language learning experience of child FL learners. In the following pages, these questions will be succinctly brought together to rough out some of the limitations that still linger on contemporary research.

First, we know very little about the impact that child oral interaction has on their written product. In this sense, collaborative writing has proven valuable in providing learners with opportunities to scaffold each other, co-construct meaning and thus collaborate in the solution of their language-related problems in ways that can facilitate L2 learning. In this manner, by taking advantage of the opportunities these communicative tasks provide, YLs can consolidate their emerging IL. The blending of both oral communication and writing throughout the text composing process is believed to offer not only the advantages provided by writing tasks, but also the advantages provided by speaking tasks. For this reason, researchers should also work on designing studies that include a written component.

Second, future studies would also need to shed light on the debate surrounding the language learning potential of WCF. Although a handful of studies has researched the effects of several WCF techniques on writing tasks with adult or adolescent learners, research on how primary school-aged children write in collaboration and benefit from WCF is scarce in the EFL context and so would seem to merit further inquiry. Besides, there is a need to explore alternative WCF techniques such as models in order to (i) shed light on the inconclusive findings on EC; (ii) analyze the benefits of those alternative techniques for YLs; and (iii) examine the characteristics of their interaction when dealing with these types of feedback. All things considered, this oral-written connection alongside the provision of feedback can together comprise a highly valuable tool to drive the children's L2 development forward.

Third, the studies conducted so far on children's writing and feedback provision concur that models play a crucial role in promoting the noticing of lexis, ideational content and chunks of language, helping children fill holes and gaps in their existing L2 knowledge as well as improving the quality and efficiency of their written texts. Nonetheless, models do not seem to be so effective in drawing learners' attention to grammatical features or in guaranteeing the successful incorporation of previously noticed ideas and expressions into subsequent written output. This happens to be particularly the case for children at beginner levels who notice surface-level differences between the model and the drafts but are not able to fully grasp the meanings of the FNs (Coyle & Roca de Larios, 2014; Coyle & Roca de Larios, 2020) or to use them accurately to improve their written texts (Cánovas et al, 2015; Coyle & Roca de Larios, 2014). In order to overcome these difficulties and limitations, the literature suggests some kind of pedagogical intervention such as (i) pre-task instruction and/or (ii) input enhancement to promote noticing and metalinguistic reflection as well as to help children widen the scope toward other linguistic aspects different from lexis, and (iii) collaborative reflection to encourage the understanding and processing of input/WCF through knowledge sharing. These are issues which also need to be at the forefront of the research agenda on children and that the present study will attempt to explore.

Fourth, the potential effect of feedback on L2 development has been evidenced to be arbitrated by intervening learner internal and external factors that affect how learners process feedback, the extent to which they notice gaps and their capacity to exploit the opportunities made available to them. In this regard, an important question which needs to be answered is whether cognitive characteristics such as proficiency level, working memory or developmental readiness among others might be mitigated by means of instruction (Cánovas Guirao, 2017). On the other hand, studies about children's attitudes to WCF are scarce, so considerably more work will need to be done to determine whether affective factors such as motivation will have an effect on their performance, whether or not their motivation changes throughout the task, and whether or not motivation is task-dependent (García Mayo, 2018). As mentioned above, given that motivation seems to be fundamental for the students' consolidation of their new knowledge (Bitchener & Storch, 2016) and that children's motivation soars and decreases during their school life (Kopinska & Azkarai, 2020), more emphasis should be put on gathering learners' beliefs and opinions on the specific tasks they perform in

their language classes in order to gain more insights into the motivational processes occurring in the L2 classroom. We also find this call by researchers within the sociocultural perspective, when researchers criticize the studies on WCF for being decontextualized, and not providing information about the context, teachers or learners' goals and beliefs that may ultimately exert an influence on the provision of and response to WCF.

Fifth, one of the shortcomings of research with school learners is that sample sizes are usually too small to be extrapolated to a wider population, even more so when the already low number of participants has to be cut in half to form pairs for collaborative work. Actually, all the studies which have explored the use of models with children have been carried out with a sample size which is not large enough to offer reliable results, thus making generalizations difficult. It is therefore clear that there is a real need for research to conduct studies on models with larger samples of participants. By enlarging the sample size, the present dissertation aims to yield more robust findings which corroborate (or not) the results reported on the use of models with YLs.

Last, but not least, firm causal evidence on the long-term positive impacts of writing practice and feedback is yet unavailable (Manchón, 2011). Although it is true that the work by Coyle et al. (2018) adds a longitudinal dimension to previous studies, they could not provide statistical analysis given the small number of participants and therefore any claims about the lasting benefits of modelling on children's L2 development may lose power. On the other hand, the results of previous studies using only one-shot feedback treatment can only be interpreted as evidence of learners' uptake rather than acquisition (Reinders, 2009). Therefore, in order to analyze feedback for acquisition, we would need a study of children's performance over longer periods of time that would enable the tracking of changes in their written production. To this end, YL's written output would have to be sequenced in different cycles of writing and feedback provision to find proof of developmental progress (Cánovas Guirao, 2017). Following Bitchener and Storch (2016) and Ortega (2012), this progress may be tracked down in the inconsistent, nonlinear and irregular use of the new L2 forms by the children. In doing so, the present study aims to contribute to increasing the knowledge about the effects of models on YLs and bring to light an uncharted research niche

regarding the lasting effects of model revision. Thereby, this dissertation breaks new ground to the exploration of the language learning potential of WCF.

Certainly, there are many issues of child L2 acquisition that would benefit from further exploration. In the present study, we hope to shed some light on those research gaps described above.

In particular, this study aims to:

(i) Obtain empirical evidence on the impact of models as a form of feedback on the final product written in collaboration.
(ii) Be able to separate the effect of the model texts from the mere task-repetition effect.
(iii) Obtain evidence on whether or not long-term engagement with writing practice and feedback can bring about learning.
(iv) Obtain evidence on the effect of input enhancement on the students' use of third person possessive pronouns.
(v) Measure student motivation throughout the process

The current chapter will introduce the research questions and their corresponding hypotheses. We have grouped the five research questions into three different modules as we consider that the reporting of the data and multiple findings would be more reader-friendly if we classify them into three different sections. The participants and the research context will also be described, along with the materials used in this study and the procedure to carry it out. Finally, the guidelines followed to codify and analyze the data will be provided.

5.2. Research questions and hypotheses

On the basis of previous research conducted on the use of models as a written feedback technique with young EFL learners, the present study addresses the following research questions and entertains their corresponding hypotheses:

Module 1: Oral-written connection

1. What linguistic features do young EFL children focus on when dealing with a task using model texts?

a) What features of language do they notice when composing a narrative text?
b) What do they notice when comparing their written texts with model texts or when correcting themselves?
c) Are there any across-cycle differences within the three groups (CG, TG and long-term treatment group (LTG)) regarding frequency, type and outcome of the episodes produced in oral interaction?

2. How is the children's reported noticing at the composition and comparison stages related to their revised texts? (Impact of oral production on written production).

Module 2: The effects of models on the children's written production

3. Do models help improve the written production of primary EFL students in the short and long run?

4. Does input enhancement play a role in the children's noticing and improvement of third person singular possessives?

Module 3: Learners' attitudes and motivation

5. What is the attitude of students toward the three feedback conditions? Does it change throughout the treatment?

1. Hypotheses regarding young EFL learners' focus when:

a) Composing a narrative text

Except for the two studies by Luquin and García Mayo (2020, 2021), who observed that grammatical features accounted for the great majority of the linguistic problems identified at Stage 1 followed by mechanics and lexis, the literature has documented that most of the problematic features that children notice while writing a text are lexical (Cánovas Guirao et al., 2015; Coyle & Roca de Larios, 2014; Lázaro-Ibarrola, 2021) The reasons given for this predominance of lexical concerns include the meaning-focused nature of the task, the type of teaching the children receive, and the learners' low proficiency level, which could have affected their dependence on vocabulary to convey basic messages. What is more, these results are in agreement with those obtained in research conducted with adult participants (Abe, 2008; García Mayo & Loidi Labandibar, 2017; Hanaoka, 2006a, 2006b, 2007; Hanaoka & Izumi, 2012; Kang, 2020; Martínez Esteban & Roca de Larios, 2010; Montealegre Ramón, 2019; Yang & Zhang, 2010).

- Taking this evidence into account, and given that part of the results obtained in the present dissertation were reported in Luquin and García Mayo (2021), we expect that the pairs of EFL children participating in our study will focus mainly on form, but also on mechanics and lexis at this stage of the writing process.

b) Comparing their written texts with model texts or when correcting themselves

Research has shown that what learners notice in the composing stage has a notable impact on what they pay attention to in the comparison stage. Therefore, it comes as no surprise that the linguistic aspects that children notice the most when feedback has been provided in the form of a model text are also of a lexical nature (Cánovas Guirao et al., 2015; Coyle & Roca de Larios, 2014, 2020; Lázaro-Ibarrola, 2021) and, unexpectedly, this was also the case in Luquin and García Mayo (2020, 2021). As VanPatten (2004) explains, children at this age and proficiency level tend to process input for meaning before they process it for form, thus paying less attention to formal aspects they consider dispensable. In addition, during feedback processing, children have to pay attention to meaning and form, notice gaps and holes in their IL, search the model for

possible solutions to their initial problems, and use that information to test hypotheses about the TL and formulate new ones. This cognitive overload is believed to push them to turn their attention to semantic and contextual clues when analyzing the model, thus pushing grammar into the background (Coyle & Roca de Larios, 2020).

To the best of our knowledge, apart from the work by Luquin and García Mayo (2020, 2021), there are no other studies on children that include a CG which self-correct their own texts. While Luquin and García Mayo (2020) observed that the children in this feedback condition spent their time identifying missing information in their writing and upgrading their texts with new ideational content, Luquin and García Mayo (2021) found that, not having any type of native output to compare with, their participants attended mostly to form and mechanics. As for studies with adult learners, Kang (2020) included a self-correction group, but did not analyze the students' foci while editing their own texts.

- Accordingly, we believe that lexical features will be the learners' priority when it comes to comparing the model with their own drafts.
- As for the children in the self-correction group, based on the work by Luquin and García Mayo (2020, 2021), we could expect either a greater concern for content or for form and/or mechanics.

c) Hypothesis regarding across-cycle differences with respect to the episodes produced in oral interaction

Up to now, only the study by Coyle et al. (2018) has made use of a longitudinal design to explore the lasting effects of model revisions in two multi-stage writing cycles. However, the authors' investigation documents the follow up of the dyads' trajectories instead of making a comparison between the number, type and/or outcome of the different episodes produced throughout the cycles.

- Thus, we have no information about differences in the amount, nature and accuracy of the episodes generated between cycles, and no study on models has included a third group to tease out task-repetition effects. However, while we would expect the TG and the LTG to behave in the same manner in the first cycle, we anticipate the LTG to produce a higher number and possibly a wider

spectrum of (correct) episodes than both the CG and the TG, given its previous exposure to models.

2. Impact of oral production on written production

When looking at the incorporations and their traceability to Stages 1 and 2, the existing studies coincide in that the changes made by learners in their revised texts are mostly lexical and content-related (Cánovas Guirao et al., 2015; Coyle & Roca de Larios, 2014; Lázaro-Ibarrola, 2021; Luquin & García Mayo, 2020), and that most of them either had not been previously reported (Coyle & Roca de Larios, 2014) or came from Stage 2 or from Stages 1 and 2 (Cánovas Guirao et al., 2015; Coyle & Roca de Larios, 2014; Lázaro-Ibarrola, 2021; Luquin & García Mayo, 2021).

Concerning the potential effects of self-correction on learners' writing, Luquin and García Mayo (2020) observed that the dyads who had edited their own texts incorporated very few corrections into the revision, and these changes were related to all the categories established (lexis, grammar, mechanics and content). Likewise, Kang (2020) observed post-test improvements in the CG, and the changes made were also equally distributed across the categories (content, organization, grammar and vocabulary). Luquin and García Mayo (2021) found that, at Stage 3, most of the features incorporated by the CG were lexical and content-related in nature (form was also very close). None of the three studies includes the tracking of these incorporations.

Hanaoka's (2006a, 2007) studies with adults included a two-month delayed post-test and revealed that the participants did not retain much of what they noticed over the course of two months, but the incorporations were mostly related to Stages 1 and 2. As for YLs, only the work by Luquin and García Mayo (2021) considered the impact of children's noticing on a delayed post-test. The authors observed that at Stage 4, the features incorporated by the CG were related to form and mechanics, while the TG incorporated mainly formal and discursive features, but also content-related and lexical items.

- Thus, as the literature suggests, it is probable that the learners in the model groups will incorporate mostly lexical features into subsequent revisions, and we

anticipate that most of these changes will be traceable to the model or will not be previously reported.

- The self-correction group is expected to incorporate few of their self-corrections into the first and second revisions. As for the nature of these incorporations, results differed, we have no basis to establish a prediction on this issue.

3. Hypothesis regarding the short- and long-term effects of model texts on the children's writing performance

Regarding short-term effects, Lázaro-Ibarrola (2021) did not find any statistically significant differences between the two drafts in the analyses carried out for CAF. However, the author reports a significant improvement in the second draft when evaluated holistically. Conversely, the global scores obtained in Villarreal and Lázaro-Ibarrola (submitted) did not reveal statistically significant differences, but the authors observed a significant increase in complexity in the model group. As for Coyle and Roca de Larios (2014), gains in the linguistic acceptability and comprehensibility of the children's revised texts showed an advantage for EC over models.

In their longitudinal study, Coyle et al. (2018) did observe an improvement in the children's final texts. The authors explain that the trajectories which had a positive effect on the children's L2 knowledge mostly involved the semantic processing of lexis and lexical phrases. The children also occasionally paid attention to and processed the models at syntactic (3rd person singular morpheme, the *-ing* morpheme, articles and personal pronouns) and discursive (cohesive elements) levels.

- Although more information is needed, in light of the results obtained by Coyle and Roca de Larios (2014), Lázaro-Ibarrola (2021) and Villarreal and Lázaro-Ibarrola (submitted), the likeliness that models have an impact on the children's writing skills in the short run seems non-existent.
- Notwithstanding the scarce information we have on the long-term effects of models on the writing performance of EFL children, we may speculate that we will find evidence of progress in the written texts, at least with respect to the incorporation of new vocabulary and expressions.

- The CG is also likely to enhance the quality of their texts given the results obtained in one-shot studies. Nevertheless, it is important to mention that there is always a risk of task procedural repetition effects (in both groups) which cannot be overlooked.

4. Hypothesis regarding the benefits of input enhancement in the children's noticing of third person possessives

As mentioned in the section on FonF (see section 2.2.2.2), highlighted saliency of the L2 forms in the textual input seems to cause noticing of the targeted L2 forms, leading to cognitive processes that benefit L2 learning (Meguro, 2019; Sharwood-Smith, 1991). Nevertheless, R. Ellis (2016) cautions that noticing is dependent on such factors as the nature of the target form itself, the student's knowledge of that form or his/her proficiency level.

- Taking all these issues into account, we may speculate that text enhancement is likely to have a positive effect on the use of possessive pronouns given their perceptiveness and the learners' familiarity with them. As we are dealing with low-proficiency students, it could also be the case that our participants find it more difficult to simultaneously grasp the meaning of the text and consciously attend to linguistic form, thus prioritizing meaning over form.

5. Hypothesis regarding the children's attitude toward the feedback conditions

Studies on the use of models which have also considered learners' attitudes towards this feedback technique and its impact on their writing performance remain scarce. Still, we can find some information on this matter in Hanaoka (2007) and Yang and Zhang (2010) who found high levels of motivation in their undergraduate participants. As for adolescent learners, García Mayo and Loidi Labandibar (2017) read through the students' written comments and observed that although they found the models quite useful, they did not really enjoy them. Furthermore, the authors reported that highly motivated students tended to incorporate more features in subsequent revisions. Kang's (2020) participants also found the models 'somewhat helpful' in improving their

writing, but the author did not provide any information regarding their motivation toward receiving the models. We know, however, that some students preferred to receive explicit corrections. Regarding YLs, to our knowledge, only the studies by Lázaro-Ibarrola and Villarreal (2021) and Villarreal and Lázaro-Ibarrola (submitted) have captured children's task motivation. In both studies, results showed that motivation ratings were high in general and evidenced a significant upward shift post-task. These scores, however, exhibited a drop in the model group at pre-task on the post-test while it was maintained in the TRG and CG, respectively. As for the motives, in line with Imaz Agirre and García Mayo (2020) and Kopinska and Azkarai (2020), collaboration was the main reason the children gave to justify their positive motivational disposition.

- In line with the research presented above we can make some predictions. In relation to motivational variation at pre- and post-task, a positive upward trend in the three groups is expected at post-task (Imaz Agirre & García Mayo, 2020; Kopinska & Azkarai, 2020; Lázaro-Ibarrola & Villarreal, 2021; Villarreal & Lázaro-Ibarrola (submitted)). With respect to the motivation ratings across cycles, although it is still unknown in which direction the task motivation of YLs will fluctuate (if it does) in studies on models, the longitudinal work by Kopinska and Azkarai (2020) on dictogloss tasks revealed that the children's disposition was highly positive and seemed to consolidate with time. Therefore, we could venture to hypothesize that the motivation of the children in both treatment groups will be maintained or increased.
- It is possible that such factors as collaboration and age provide a good basis for a positive attitude and, as a result, for a good performance. Interaction in writing is still a relatively new concept for teachers, so students who are not used to working jointly may find the task amusing and may feel eager to receive the feedback. In contrast with García Mayo and Loidi Labandibar's (2017) adolescent participants, children are not yet undergoing emotional changes and therefore they may be more positive toward the task. Thus, they are expected to be highly motivated.
- Although we do not have information about children's (or adults') attitudes toward self-correction, we know that learners are usually lost when it comes to correcting their own texts, and this happens especially with YLs (Luquin & García Mayo, 2020). Therefore, we may speculate that not only leaving

correction up to children, but also having to repeat this task without serving a real purpose or receiving any type of feedback could decrease their motivation toward the task considerably (Nitta & Baba, 2014).

5.3. Participants and setting

The participants in this study were 60 11- to 12-year-old bilingual (Basque and Spanish) children (27 boys and 33 girls) forming a total of 30 pairs from three EFL classes in a semi-private school (partially funded by the local government) in Northern Spain. They had been studying English for about 7 years and received 3.5 hours of English classes and 3.5 hours of content-based instruction (during their science, art and music courses) through English per week. These classes follow a communicative approach based on the integration of all skills and on promoting interaction in the L2 by making an extensive use of role-plays, games, songs, multimedia presentations, realia or conversation activities. Authentic materials and resources created for native speakers of the TL such as movies, TV shows, picture books, folk songs, story books and advertisements are also used in the classroom. Thanks to the CLIL approach, which entails a more frequent use of the FL in the classroom, the learners in the study were used to and felt comfortable using the TL with the teacher and with their classmates. The children's proficiency level in the TL was an A2 level in the Common European Framework of Reference for Languages as attested by the Flyers test (Cambridge Young Learners English Assessment, 2018), administered and corrected by the researcher (average score of 6.09 out of 10). The test consisted of a listening (matching information with pictures and/or names, fill-in-the-gap exercises or coloring parts of a picture), a reading and a writing activity (choosing the correct words according to their definition, selecting the best answer to a question, identifying specific information in a text, filling in the blacks with the most suitable word and answering comprehension questions). Before the data were collected, the children also filled in a language background questionnaire (see Appendix 1 and the questionnaire's translation below) in order to discard potential outliers. In addition, informed consent was obtained from the school principal and the participants' parents for the children to be recorded and take part in the research (see Appendix 2 and its translation below)

The three EFL classes were randomly assigned to a CG (n=18/9 pairs), a TG (n=22/11 pairs) and an LTG (n=20/10 pairs). Within each group, the children were placed in proficiency-matched pairs according to the scores they achieved in the Flyers placement test, but we also asked the teacher to make the necessary changes to the pairing taking into account their ability to work well together. The type of relationship students form was deemed relevant in order to avoid an unequal contribution of each dyad member to the writing task, an undesirable situation that usually emerges in mixed-proficiency pairs. Through this level-match pairing, we expected to guarantee a good level of mutuality (the level of learners' engagement with each other's contributions) and equality (the level of learners' contributions to the task) (Storch, 2002), thus avoiding potential asymmetrical relationships which could negatively affect their interaction (Storch, 2016). In addition, the high-high proficiency learners would benefit from each other's expertise, and the low-low proficiency learners would more likely feel less intimidated and participate more in resolving LREs. Table 2 shows the details of the participants in this study.

Table 2. Details of the participants in the study

CG	TG	LTG
n=18 (9♂, 9♀)	n=22 (10♂, 12♀)	n=20 (8♂, 12♀)

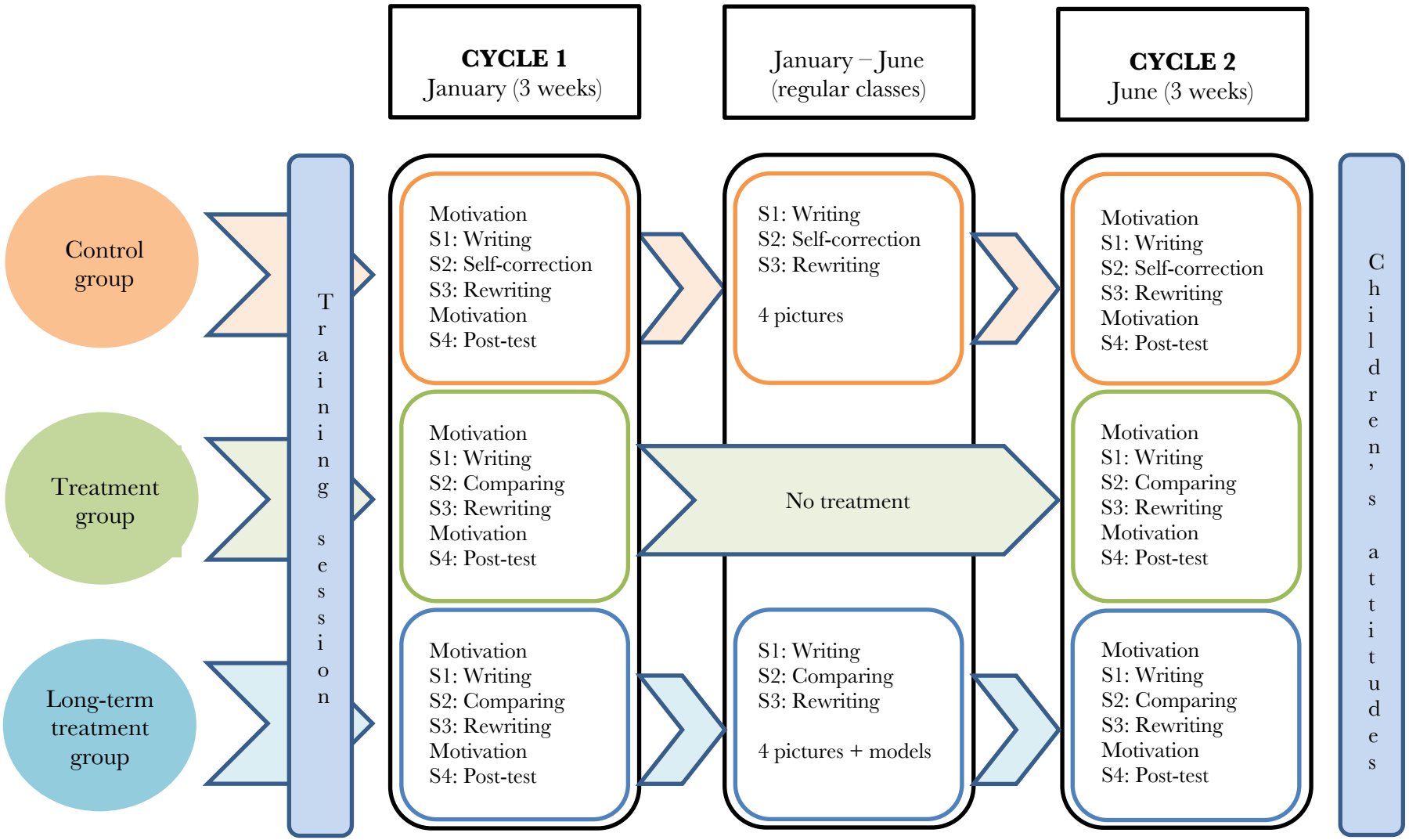
5.4. Research design

The present study was designed with a view to analyzing the effects of models over a sustained period of time. It was therefore expected that looking at the progress of the three groups mentioned above under different conditions and comparing their written production across four different time periods (their original drafts and subsequent revisions in two writing cycles) might be an appropriate methodology for the aim of the research. Additionally, this study would allow us to integrate quantitative and qualitative data for a richer and more comprehensive understanding of the research topic and to know more about the language learning potential of collaborative writing and the use of models as an alternative form of feedback. To this end, we engaged the three groups in a task consisting of two four-stage writing cycles of 3 weeks each (writing

initial text, comparing/self-editing, rewriting the text again, and delayed post-test) and separated by four months. Moreover, the CG and the LTG continued working with their teachers on their corresponding feedback technique during this four-month period between both cycles. Data were collected over a period of six months (from January to June) as shown in Figure 2 (more detailed information in Appendix 3).

The children were divided into three groups: (i) the TG, who received treatment only in January and in June; (ii) the LTG, who received treatment from January to June and whose purpose was to isolate any potential feedback effects resulting from the long-term treatment; and (iii) the CG, who did not benefit from any treatment nor did they receive feedback on their written production, but self-corrected their texts from January to June. In this regard, it is important to mention that, although depriving a group of children of a potentially valuable feedback treatment might be regarded as an ethically contentious issue, we must say that the inclusion of a CG enabled us to provide clearer evidence-based research on the effects of models on L2 learners' noticing and writing performance.

Figure 2. Research design



As explained in the first part of the present study (see section 2.3), the decision to make writing collaborative was motivated by available research with adults and children which has shown that collaborative writing activities can engage students in deliberations about language, and that such deliberations lead to progress in L2 development and benefits for writing (e.g., Calzada & García Mayo, 2020a, 2020b; Storch, 2005; Wigglesworth & Storch, 2012). It is precisely when the students share the responsibility of the oral or written product of a task that they collaborate in the solution of the problems they face. As a result, learners produce jointly discussed output that enables them to solve problems and test hypotheses on which they receive feedback (Storch, 2016). All things considered, the children in the present study were asked to work in dyads, and to try to identify, find solutions to, comment on and discuss any problems they might experience during the different tasks.

In view of the lexical gains that models seem to offer to the detriment of grammar, we decided to bring the children closer to formal aspects throughout the processes of writing and comparison with the model. As Loewen and Sato (2018, p.308) observe:

> It is perhaps unsurprising that vocabulary should figure so prominently in meaning-focused interaction, especially when learners are left to their own devices. Consequently, there may be less of a need for manipulation of lexical items before, during, and after interactive tasks, and a greater need to draw learners' attention to grammar.

Accordingly, we aimed to help the children better exploit the advantages of this feedback strategy by drawing their attention to formal features in three different ways. First, we decided that a pre-task training session could be useful not only to raise the children's awareness of a wider range of linguistic aspects, as R. Ellis (2016) points out, but also to familiarize them with a task which could become cognitively demanding for YLs (Allwright, Woodley & Allwright, 1988; Barnawi, 2010; Cánovas Guirao, 2017; Cánovas Guirao et al., 2015; Coyle et al., 2018; Luquin & García Mayo, 2020; Qi & Lapkin, 2001; Yang & Zhang, 2010). Accordingly, before the actual research started, a session was devoted to comparing a model text with some of the drafts the children had previously written in pairs. Second, results from studies on collaborative writing have shown that there is greater grammatical precision in texts written in collaboration and

that learners consider that this type of writing allows them to join forces to generate new ideas and, above all, pay attention to formal aspects of the language (Alegría de la Colina & García Mayo, 2007). Therefore, the reason behind having the children write a text collaboratively was twofold: as stated above, to enable the students to collaborate in the solution of the problems they encounter and pool their linguistic resources, and to boost the noticing of aspects other than lexis. Third, one of the issues that is capturing the researchers' attention nowadays is the use of input enhancement, which rests on the premise that highlighted saliency of the L2 forms in the textual input would cause noticing of the targeted constructions, leading to cognitive processes that benefit L2 learning (Meguro, 2019; Sharwood-Smith, 1991). The difficulties that many Spanish and bilingual (Basque/Spanish) EFL children seem to experience with distinguishing masculine from feminine possessive pronouns (see Imaz Agirre & García Mayo, 2013, 2018) led us to employ this technique in the model texts in an attempt to assist the children in relation to this matter and without making use of explicit instruction.

Another issue which we considered should be at the forefront of the present study was the qualitative analysis of children's language learning attitudes and motivation toward models and self-correction as well as of their impact on the children's written performance. To this purpose, the learners were asked to complete a motivation thermometer before and after performing the task in each cycle. Furthermore, after cycle 2, a representative sample was selected from each group to form focus groups which were interviewed to explore this issue in more depth.

Finally, the existing studies on the use of models fail to contemplate possible procedural or exact task-repetition effects on the results reported. Additionally, students' learning has been gauged by having them work on the same writing task. This practice would seem to question the authenticity of any conclusions concerning learners' mental representations of linguistic knowledge. In other words, 'unless their development is measured in a new piece of writing, it remains questionable as to whether they have understood the rules behind linguistic manifestations in the revision phase or simply reused some memorized chunks' (Yamashita, 2021, p. 77). These are the reasons why we felt that it was necessary to include a fourth stage in each of the cycles which would serve as a delayed post-test. This way, we could (i) elucidate whether any gained knowledge from the feedback is retained in the short term and under different

conditions, and (ii) tease out same task repetition effects by using a different picture prompt. Likewise, we considered it appropriate to have the CG also participate in the period between cycles. In this way, a post-hoc comparison between the CG and the TG would give us information about the extent to which task repetition has acted upon the overall results.

5.5. Data collection

Before collecting the data, the children were provided with guidance and practice in using models so that they could better exploit the feedback treatment. The data collection procedure commenced right after this training session and consisted of two four-stage writing cycles and a long-term treatment in between the cycles. All these procedures are explained in what follows.

5.5.1 Training session

In view of the failures to notice certain features in the feedback, the misunderstandings of input, or the unfamiliarity with the task reported in the studies conducted with children on modeling, more and more researchers call for extended practice on feedback comparison and processing at these early ages (Cánovas Guirao et al., 2015). In order to answer this call to the extent possible, we held a 50-minute training session to work with models and familiarize the children with them. Although the children were able to communicate in the L2, we decided to use the L1 on this occasion in order to guarantee a full understanding of the procedure. The session was devoted to comparing a model with some of the texts that the learners had previously written in pairs in response to a picture (see Appendix 4) during one of their regular EFL lessons. The model text (see Appendix 5) and the picture, which was a made up of six frames (about a girl who ruins her brother's homework by doing crafts on it), were taken from the textbook *Young learners English tests* (Cambridge University Press).

All the learners, irrespective of the groups they had been assigned to and whether or not they were going to participate in the study, benefitted from the training session, which

was carried out through whole-class discussion. Both the model text and the samples of the pairs' first writings (see Appendix 6) were shown on the school auditorium's projection screen, while the blackboard was used for explanations, diagrams and drawings. After reading the two texts, the researcher asked the children to work with their corresponding partner for a couple of minutes and try to identify all the similarities and differences between the model and their classmates' drafts. The children soon began to point to similarities and differences without providing any explanation for their answers, turning a whole-class discussion into a 'spot the difference' game. In order to avoid this situation, also observed in previous studies with children (Cánovas Guirao, 2017; Cánovas Guirao, et al., 2015; Luquin & García Mayo, 2020), the learners were also requested to give a reason for these differences. Finally, many YLs ventured to provide an explanation for every mismatch found between the texts. Nevertheless, the researcher had to intervene regularly, as they could not always provide a metalinguistic explanation, in particular with formal aspects.

The researcher also attempted to widen the scope toward other linguistic aspects different from lexis, which seems to be the overriding focus when dealing with models, as previous research has detected (Abe, 2008; Cánovas Guirao et al., 2015; Coyle & Roca de Larios, 2014; Coyle & Roca de Larios, 2020; Coyle et al., 2018; García Mayo & Loidi Labandibar, 2017; Hanaoka, 2006a, 2006b, 2007; Hanaoka & Izumi, 2012; Kang, 2020; Lázaro-Ibarrola, 2021; Luquin & García Mayo, 2020; Martínez Esteban & Roca de Larios, 2010; Montealegre Ramón, 2019; Yang & Zhang, 2010). Accordingly, we tried to make the YLs aware of (i) grammatical, (ii) lexical, (iii) spelling, (iv) content-related errors as well as of (v) stylistics and (vi) discursive devices. The decision to focus on these aspects was motivated by the categories used by Cánovas Guirao (2017) who, in turn, based her decisions on the categories proposed by Qi and Lapkin (2001) and Yang and Zhang (2010). As a result, when the participants found a difference, the researcher wrote it on the blackboard, elicited a justification from the students and completed or corrected the answer by providing a metalinguistic explanation.

Throughout the session, the researcher noticed that the learners were able to notice increasingly more differences between the original drafts and the native text, and a wider spectrum of categories.

5.5.2 Cycle 1

Once the training session was over, data collection started with the first cycle, which took place in January of the academic course 2018-2019 and over a period of three weeks. Cycle 1 consisted of four stages: Stage 1 (or writing stage), Stage 2 (or comparison/self-correction stage), Stage 3 (or rewriting stage) and Stage 4 (delayed post-test). The first stage took place on the first day (a Monday), the second stage on the following day (Tuesday), and the third stage was held one week later to eliminate the possible effects of memorization (Monday), that is, to reduce the chance of participants recalling the model text and retrieving its content. The final stage served as a delayed post-test and was performed a week after Stage 3 (Monday). In order to prevent the learners from being bothered during the sessions, the school provided three small and quite separate rooms in which the researcher and two assistants met with one dyad each. Thus, three pairs were performing the task simultaneously in three different rooms. When each pair finished the task, another pair was convened. While the dyads performed the different tasks, the rest of their classmates continued with their regular lessons. All discussions at every stage were video and audio recorded in the presence of the researcher and her assistants. Both video cameras and digital voice recorders were used in order to capture the finest nuances. It is important to clarify that guided note-taking is a common practice in studies on models given its effectiveness in promoting noticing (Abe, 2008; Cánovas Guirao et al., 2015; Coyle & Roca de Larios, 2014; García Mayo & Loidi Labandibar, 2017; Martínez Esteban & Roca de Larios, 2010). However, we discarded this technique in order to avoid a processing burden, as the participants were too young to cope with a triple task (writing/comparing, interacting and note-taking) which is physically demanding and time-consuming and might put their motivation at risk. Therefore, we considered it sufficient having the children speak about their linguistic problems or whether the alternative was appropriate or not without having to report about it.

Besides having the instructions written on the composition sheet (see Appendix 6), additional instructions were also given in English as the children were used to the language not only because of their mainstream classes but also because of the content-based program. Accordingly, the participants were encouraged to use English as much as possible in the course of their interactions, but the use of Spanish and/or Basque was

not forbidden in case the children were not able to make themselves understood. Nevertheless, the use of their L1s was practically non-existent. Regarding the time allotted, a time limit of 15 minutes was established for each task. This decision was made taking into account the average amount of time (11.8 minutes) that the dyads in the pilot study (Luquin & García Mayo, 2020) took to complete each task. Besides, 5 additional minutes were granted to fill in the motivation questionnaire (see below) at stages 1 and 3. The model text chosen for Cycle 1 was approved by the children's teachers, who considered it appropriate for their L2 proficiency level.

The output, comparison and revision tasks were expected to give the children the chance to experience language problems or even notice holes between what they wanted to say and what they were actually able to say during the writing process; to notice gaps between their IL and the TL when comparing their first draft to the model text; and to feel that they could incorporate new L2 features into subsequent revisions. These hopes were motivated by existing research which has reported the value of multi-stage tasks in eliciting those practices among L2 writers (e.g., Abe, 2008; Coyle & Roca de Larios, 2014, Yang & Zhang, 2010).

At Stage 1 (writing stage), in order to measure their motivation at pre-task, each participant completed a *Motivation thermometer* (see Appendix 7). This thermometer, adapted from Al Khalil (2016), was a questionnaire which asked them to rate how motivated they were to do the activity on a 10-point scale (where ten represents the highest motivation possible), to underline the reason(s) for the score provided and/or add one reason of their own. This activity took them about 5 minutes to complete and was carried out Spanish to ensure that the questions and the reasons provided were understood (a translation into English has been included below Appendix 7 for the reader's convenience). Afterwards, the children in the three groups were requested to write a story in pairs in response to a visual prompt ('Martine's alarm-clock', originally used by Lapkin, Swain, & Smith, 2002 in black and white) which illustrates a girl who is late for school and rushes to arrive on time (see Appendix 8). The six-frame picture story prompt was colored by the researcher in order to make story more appealing and did not contain any explanatory text that could reveal the story plot. No indications were provided concerning the use of specific vocabulary, discursive devices or verb tenses,

but the learners were left to complete the task using English as they best knew. Finally, they were invited to proofread their composition before submission to ensure that the errors were not just oversights that they could self-correct.

At Stage 2 (comparison stage), the pairs in the two treatment groups were provided with the stories they had written, the cartoon strip, and its corresponding model text (taken from Lapkin, Swain & Smith, 2002) (see Appendix 9) which had been carefully chosen taking into consideration the children's L2 level. As the children were not yet entirely familiar with the past tense, the present tense used in this first model text was maintained. Nevertheless, some features which could become challenging as they were above the children's current level of competence were also kept so that we could assess the learners' language progress. Thereby, the native texts contained familiar words (e.g., girl, bed, school, hand, alarm, etc.), but also some unknown words and expressions (e.g., fall asleep, feather, comb, tickle, etc.) that the participants had not yet seen in their L2 classes, but which were considered accessible to them.

In addition, and as indicated earlier, input enhancement was used as one of the procedures to direct the children's attention to grammatical aspects. Actually, previous studies have shown that highlighted saliency of input facilitates the noticing of the targeted L2 forms, leading to cognitive processes that benefit L2 learning (Doughty & Williams, 1998; Lee & Révész, 2018; Meguro, 2019; Sharwood-Smith, 1991). Taking into account R. Ellis' (2016) note of caution that noticing depends on the degree of saliency and on the children's familiarity with the targeted item, among other factors, third person singular possessive pronouns were chosen as a target construction. Spanish lacks gender agreement for third person singular possessives, being *su* the only form used for either of the three genders, which explains why children at these ages tend to struggle with this linguistic feature. Given that YLs usually experience problems with both masculine and feminine possessive pronouns due to their typographical similarity (*his/her*), we ensured that each model only contained same-gender pronouns to avoid confusion, and that the number of male and female main characters in all the models used in the study was balanced. As for the enhancement technique, the target construction was highlighted via underlining, on the grounds that this form of enhancement might be more effective in creating an isolation effect, thus fostering the

visual salience of the targeted grammatical construction to a higher extent than boldfacing or using different color fonts (Indrarathne & Kormos's, 2017).

At this stage, the students were instructed to comment upon and discuss all the differences and similarities they could notice between their original drafts and the native version, just as we had proceeded in the training session, but no comment was made on the underlined words. If some child asked why some words were underlined, the researcher and the assistants agreed to tell them that we did not know why. The pairs in the CG, on the other hand, were asked to self-correct their collaborative text in pencil and try to explain the changes made.

At Stage 3 (rewriting stage), each pair in the three groups was given the picture again, but they were not allowed to revise either the model or their initial drafts. Then, the children were instructed to rewrite the story and try to recall and incorporate the items noticed in the feedback the previous week. The students were not informed about this task beforehand in order to avoid memorization of the corrections. After the rewriting was completed, our participants filled in the post-task motivation questionnaire (see Appendix 10 and the questionnaire's translation below), thus enabling us to have a broader perspective of whether their rating varied.

Stage 4 (delayed post-test), held one week after Stage 3, consisted in the production of a text based on a new visual prompt which narrated a story similar to the first one and also had the same number of cartoons (see Appendix 11), so that the students could incorporate some of the features present in the first story. We intended to use a different visual stimulus, but not too different so that the tasks were isomorphic. The purpose was twofold: on the one hand, we wanted to see how much of the indirect feedback they were able to remember and retain in the short term. On the other hand, we expected to tease out same task-repetition effects from actual learning from the model to the extent possible. That way we ensured that the new linguistic knowledge that the children could have stored and retrieved was not a consequence of having performed the same task twice. To this purpose, 'One more day to school' was specifically designed for this study by an illustrator (©Israel Azpilicueta), with the intention that the children were given the opportunity to use newly learned material (e.g., the sun is rising, a girl is sleeping on her bed, have sweet dreams, feather, alarm-clock, ring, brush, comb, etc.) in

a different context. What is more, the fact that the protagonist was a girl allowed for the implementation of feminine possessive pronouns.

5.5.3 Extended treatment period

From February to May, the LTG and the CG were provided with one picture prompt per month and completed stages 1 (writing), 2 (comparison or self-correction) and 3 (rewriting) with their corresponding EFL or CLIL teachers. Therefore, the CG received no treatment and the LTG benefitted from 4 model texts. In this way, we could corroborate whether or not a long exposure to models does have a real impact on the students' written production and the development of the learner's IL. The presence of the CG would allow us to control for potential task-repetition effects, in such a way that if the CG eventually performed better than the TG, it could be ascribed to a task-repetition effect. On the other hand, if the LTG obtained better results than both the CG and the TG, we might claim that task-repetition could not be the only reason for the improved performance. The children in the TG, however, did not benefit from any treatment during this period. They attended their regular EFL and CLIL lessons as usual during the six months in which the CG and the LTG took part in their corresponding feedback approach.

All the pictures and their corresponding texts used to this effect were carefully chosen taking into account the visual stimulus and the model text provided in Cycle 2. This would allow us to see whether the children incorporate new L2 forms into the written tasks of Cycle 2, thus following up the students' progress. All the images contained six frames and most of them had to be colored by the researcher. In addition, all the model texts were approved by the teachers. They are explained as follows:

- In February, the groups worked with 'The cat has six lives' (see Appendixes 12 and 13), which told the story of an old woman's cat who climbs a tree trying to run away from a dog. The six-frame colored picture and the text were taken and adapted from Cánovas Guirao (2017). Some words and constructions were changed for more level-appropriate ones, and some others were added to the story to make it longer as well as to provide the children with (underlined)

female possessive pronouns. The story was written in the present, but we did not modify this aspect since the teaching of the past tense was scheduled for April.

- The story used in March was about a witch who turns a cat into a bat upon seeing that the cat was eating her food (see Appendixes 14 and 15). Both the picture and the model were also taken from Cánovas Guirao (2017), and some changes were made to the story so that some of the features matched those included in text used for Cycle 2. The female possessive pronouns were also underlined as usual.

- In April, the children in both groups were provided with the picture prompt 'Ssssurprise!' (taken from the digital repository of instruments and materials for research into second languages (IRIS) (iris-database.org)), a story about a man whose suitcase is stolen, and the thieves find a snake inside (see Appendixes 16 and 17). As no model text was available, we had a native speaker of English write the story. We asked her to include masculine possessive pronouns and write the text in the past tense, since the children were already working on it.

- Finally, in May, the participants' last prompt was 'The table that got smaller' (taken from the IRIS database repository), a sequence of pictures about two children who are playing ping pong but as the table is too high, they decide to saw its legs (see Appendixes 18 and 19). As happened with the previous image, a native speaker wrote the model text for us, also in the past and including masculine possessive pronouns.

5.5.4 Cycle 2

The second four-stage cycle took place immediately after the treatment period and lasted a further 3 weeks. In this cycle, which was implemented in June, the same four-stage task (writing, comparing/self-correcting and rewriting) as in cycle 1 was carried out by the children in the three groups and the pre- and post-task motivation questionnaires were also administered. The exact same procedure used for the first cycle was also used for the second, the only difference being that different six-frame picture story prompts were used: 'The Scientist' (see Appendixes 20 and 21) was about a

scientist who drinks a potion and turns into a cat. The prompt and the text were taken from Cánovas Guirao (2017), but the picture was colored to make the task more motivating for the children, and the masculine possessive pronouns in the text were enhanced. 'The bat-scientist' (see Appendix 22), specifically designed by the illustrator (©Israel Azpilicueta) for this second delayed post-test, told a similar story, but this time the scientist turns into a bat. In this cycle, it was expected that any potential gains obtained from the treatment provided from February to May could be observed. Put it differently, we wanted to verify whether the participants were able to put into practice their experience with models to improve their noticing and ultimately the overall quality and efficiency of their written texts. As indicated in the previous section, the model text used to this purpose included many words and expressions that the children had come across in the models provided the previous four months (e.g., laboratory, bark, conduct experiments, bat, suddenly, immediately, loud noise, bright flash of light, etc.). Accordingly, the children in the LTG were expected to use these features (even the past tense) to a higher extent than their counterparts in the CG and the TG.

5.5.5 Focus group interviews

One day after Cycle 2 was completed, we gathered the information about the children's attitudes toward models and self-correction. This gathering of data comprised two parts: an individual questionnaire and focus group interviews. For the completion of both activities, six students from each group were selected randomly to form the focus groups, but we made sure that none of their partners was chosen in order to avoid compromised responses. One at a time, the three focus groups completed the activities in one of the rooms provided for the project. Spanish was used in both the survey and the interviews to ensure that the questions were fully understood and that the participants felt comfortable answering. Specifically, the process was as follows.

First, the participants were asked to complete an anonymous questionnaire individually (see Appendix 23 and its translated version) where they had to give their opinion on the tasks performed. The survey was based on Loidi Labandibar (2016), and its objective was to gather some qualitative data about the children's attitudes toward comparing their text with a model or toward self-correction. This questionnaire allowed us to uncover the learners' thoughts and to assess the possibility of the children's

performance being influenced by affective factors. As a couple of questions changed depending on the group the survey was addressed to (CG or TG/LTG), two different models were used. The questionnaire contained two items eliciting specific information about the writing and comparison or self-correction stages and an open item which asked them to provide any further comment or suggestion.

Once the children finished completing the questionnaire, the researcher moved on to interview each focus group in an effort to explore the above issues in more depth (see Appendix 24 and its translated version). In fact, according to Lewis (1992), group interviews may trigger different answers in YLs compared with individual interviews and can also generate a greater range of answers. The interviews were video and audio recorded and lasted approximately 20 minutes each. The children were asked to take turns to speak and to be honest in their answers as their identities were not going to be revealed. The interview questions were designed with the purpose of obtaining more complete answers than those provided in the questionnaire and of addressing aspects which may have not been covered in the motivation thermometer questionnaire or in the previous survey. In particular, they were asked questions such as how they would like to be corrected, how they thought the activities were going to be like, in which ways they think their English has improved or whether they liked working with their partner.

5.6. Data analysis

The dataset for this study consisted of the following:

1) 60 joint-written compositions and 60 audio and video recordings with all the participants' discussions about their first writing (Stage 1 - Cycle 1 and Cycle 2)

2) 60 audio and video recordings of the pairs' conversations during the comparison of their written output with the model text or of the students' self-correction (Stage 2 - Cycle 1 and Cycle 2)

3) 60 sets of revised texts and 60 audio and video recordings with the children's discussion of the revised story (Stage 3 - Cycle 1 and Cycle 2)

4) 60 delayed post-tests and 60 sets of audio and video recordings of the pairs' conversations during the post-test (Stage 4 – Cycle 1 and Cycle 2)

5) 120 individual pre- and post-task motivation questionnaires (Stage 1 and Stage 3 – Cycle 1 and Cycle 2)

6) 18 surveys and 3 recorded interviews on the children's attitudes

In order to answer the two first research questions which are concerned with the children's noticing, the data were analyzed within each individual stage separately. All the collaborative dialogues recorded (60 hours, 20 minutes and 54 seconds) were transcribed for analysis using ordinary orthographic conventions, and the resulting transcripts were coded using NVivo (https://www.qsrinternational.com/nvivo-qualitative-data-analysis-software/home), a widely used software for organizing and analyzing data. The thesis supervisor coded 48 of the participants' interactions, which constituted 20% of recorded pair talk. The inter-rater reliability was found to be 96% after one round of discussion, and any remaining discrepancies were solved after considering each case individually. Statistical analyses were conducted using SPSS (version 27.0) and JASP Team 2021 (version 0.16) software systems. For the third research question, which deals with the long-term effects of modeling on the participants' writing performance, the original and revised texts produced in both cycles were transcribed and type of clause, complexity, accuracy and fluency (CAF) as well as holistic measures were used in search for evidence of any gains in the students' written output. To answer the question about input enhancement, the first (Stage 1, Cycle 1) and last (Stage 4, Cycle 2) writings were compared in terms of correct use of third person possessive pronouns. Finally, in order to answer the question on the children's attitudes, the surveys and the interviews were also transcribed and coded in NVivo by running word frequency queries and creating theme nodes. The transcriptions of the students' written output, surveys and interviews were analyzed qualitatively. In the following pages, a detailed description of how the data was analyzed and codified is provided.

5.6.1 Analysis of noticing and the oral-written connection

In order to answer research questions 1 and 2, which deal with the children's noticing and its impact on subsequent revisions, we have taken the following steps.

To start with, 'noticing' was operationalized as the children's attention to surface features in their written texts and in the feedback. Nevertheless, following Yang and Zhang (2010), we decided to isolate the noticing of linguistic problems from the noticing of those cases which exhibited children's problems with features which did not appertain to language itself, but rather to making sense of the sequence of pictures or generating content ideas. The detection of these problems was accomplished with the help of the category 'content-related episode' (CRE), which has been defined by Yang and Zhang (2010) as 'any segment in pair talk concerning content clarification or idea generation other than language-related problems' (p. 469).

On the other hand, each linguistic problem identified in the transcribed dialogues was classified as an LRE. Qi and Lapkin's (2001, p. 287) definition for LRE is provided again below for the reader's convenience:

> a segment of the protocol in which a learner noticed a language-related problem he/she encountered while comparing his/her text to a model and addressed it either by accepting the model and providing a reason, or only noticing the difference without giving a reason.

In addition, also following the procedure in Qi and Lapkin (2001), we did not tally as LREs verbalizations of parts of the text as they were being written, since they constituted no evidence of the existence of cognitive processes intervening in the resolution of a linguistic problem. With Qi and Lapkin (2001), we viewed these verbalizations as an automatic act of writing. Nevertheless, we did count LREs which were not verbalized (e.g., self-correction) but captured in the videotape as LREs. For instance, at Stage 1, while proofreading her first writing, CHILD13 crossed out the word 'have' in 'Sarah have' and wrote 'has' above it without verbalizing the process.

Following previous research (Coyle & Roca de Larios, 2014; García Mayo & Loidi Labandíbar, 2017; Hanaoka, 2007) and in order to make it clear at which stage the episodes were generated, both CREs and LREs were further classified as follows: problematic features noticed (PFNs) by the children at Stage 1, that is, any content and language-related problems that the participants identified while writing the first text; features noticed (FNs) from comparison with the feedback at Stage 2, and both features

noticed and incorporated (FNIs) and features unnoticed and incorporated (FUI) into both the rewriting (Stage 3) and the delayed post-test (Stage 4), depending on whether or not the items had been previously perceived at stages 1 and/or 2.

That said, in the first place, the data were analyzed within each stage separately before proceeding with the across-stage analysis. In all four stages, the transcribed dialogues were segmented into episodes and these episodes were classified according to the focus of attention. After the first reading, segmentation and classification of the data, the coding categories were further refined.

The focus of attention

The dialogue protocols were coded according to whether the aspects attended to were content- (CRE) or language-related (LRE). LREs were in turn classified taking into account not only the linguistic features our child participants mostly focused on, but also the most frequently used categories and classifications in the studies with models (see Appendix 25). In consonance with these schemes, five main categories were distinguished:

- *Mechanics* (M-LRE): Episodes in which the children focused on issues related to spelling, punctuation and pronunciation. A unique category for mechanical aspects was created out of 'form' (see below) in view of the amount of spelling and punctuation features the children usually focus on.
- *Lexis* (L-LRE): This category includes LREs that revolve around the meaning of verbs, adjectives, adverbs, prepositions[2], nouns, pronouns, noun phrases (NP) and verb phrases (VP).
- *Form* (F-LRE): Episodes in which the children focused on issues related to verb form (agreement), verb tense, subject, object, plural, possessive marker, possessive pronoun, comparative or superlative, pronoun, adverb, word order, article and preposition[3].

[2] Discussions about the meaning of prepositions.

[3] Discussions about the use of prepositions.

- *Discourse* (D-LRE): Those features related to logical sequencing such as coherence and cohesion, the use of paragraphs and stylistics (the aesthetic part of the text).
- *Other*: Other aspects that did not fall into any of the categories mentioned above. In this category, we may find examples where the children realize that something is missing after revising the whole text or comments regarding the learners' assessment of the overall quality of the models or of their own texts.

In order to better understand the classification of the aforementioned categories and subcategories of episodes, Table 3 illustrates the different categories with examples.

Table 3. Classification of episodes

Categorization of episodes		
CREs	**Examples**	
Content	**Content**	Pair 4, CG, Stage 1, Cycle 1
	CHI8: And go walk to school, ok CHI7: But igual no es para ir al cole [But maybe she doesn't get up to go to school] CHI8: Bueno… [Well…] CHI7: Bueno, sí, porque quién se levanta a las 7 de la mañana…? [Well, yes, because who gets up at 7 in the morning…?]	
LREs		
Mechanics	**M-LRE (spelling)**	Pair 21, LTG, Stage 4, Cycle 1
	CHI41: How can you say… o sea [I mean], how do you write 'hair'? CHI42: Again?! CHI41: Yes, I don't know! CHI42: (Writes 'hair'). Study!	
	M-LRE (punctuation)	Pair 24, LTG, Stage 1, Cycle 1
	CHI47: In her bed CHI48: Punto [Period]. Ok CHI47: Then… CHI48: Then… CHI47: Comma, no? Or point [period], no CHI48: Point [period], I put dot (referring to a period) CHI47: Ok	
	M-LRE (pronunciation)	Pair 10, TG, Stage 1, Cycle 2
	CHI19: Very… And he transformation (Spanish diction [tɾãnsfoɾˈmatɪɔn]) and the cat…	

	CHI20: And he… CHI19: Then, he transformation ([tɾãnsfoɾˈmatɪɔn]) CHI20: Transformation (corrects pronunciation [ˌtrænsfərˈmeɪʃən]) CHI19: Ah, sí, transformation… (pronounced correctly) in cat, and…	
Lexis	**L-LRE (verb)**	Pair 2, TG, Stage 3, Cycle 1
	CHI22: Turn off the clock and she… 'volver'? ['return'?] CHI23: Return CHI22: Return. She return…	
	L-LRE (adjective)	Pair 10, TG, Stage 1, Cycle 2
	CHI19: And he was very… malo [bad]… yes, very, very… cómo se dice? [how do you say it?] CHI20: Qué? [What?] CHI19: 'Malo'. Very horror CHI20: Very horror person. And he has eh… a potion	
	L-LRE (adverb)	Pair 12, TG, Stage 1, Cycle 2
	CHI23: Suddenly, he finish it CHI24: He finish it. But no, is… 'suddenly' no, eh… It's not 'suddenly'. It's eh… 'soon', and… and he…	
	L-LRE (preposition)	Pair 4, CG, Stage 3, Cycle 1
	CHI7: And the clock is behind the bed CHI8: No, 'behind' es 'detrás' ['Behind' means 'behind'] CHI7: No, es 'al lado' [It means 'beside'] CHI8: No, yo creo que no, ¿eh? [I think you're wrong] CHI7: 'Beside' es 'además' ['Beside' means 'besides'] CHI8: Yes. Vale [Ok] CHI7: And the clock is behind the bed	

L-LRE (noun) Pair 13, TG, Stage 1, Cycle 2

CHI25: Strange… pass… (meaning ‘happen’)
CHI26: Something strange pass… Y cómo se dice ‘garganta’? Bueno, in his neck, yo qué sé [And how do you say ‘throat’? Well, in his neck, I don’t know]

L-LRE (pronoun) Pair 5, CG, Stage 3, Cycle 1

CHI10: Sí, who’s name (they laugh). Whose name, whose name, whose name. Venga! [Come on!] (Speaks unintelligibly). O whose… Es que creo que es lo mismo. Creo que es abreviado, eh? [I think they mean the same. I think it is contracted]
CHI9: Whose name
CHI10: ‘Who’s’ es ‘quién’ [‘Who’s’ means ‘who’]
CHI9: Qué bien [Great]
CHI10: ‘Who’s’ is ‘quién’
CHI9: Que me da… [I don’t care…]
CHI10: y ‘whose’ es ‘cómo se llama’ [and ‘whose’ means ‘what’s her name’]
CHI9: Sí [Yes]
CHI10: Ok (laughs). ‘Who’ es quien y ‘whose’ es como… [‘Who’ means ‘who’ and ‘whose’ means ‘what’s….] Who’s go to the party? (laughs). Martin…

L-LRE (NP) Pair 4, CG, Stage 3, Cycle 2

CHI7: She… he… start with eh…
CHI8: Dolor de garganta [sore throat]
CHI7: Neck ache. Neck ache
CHI8: (Shrugs her shoulders). No sé [I don’t know]. Ah! No, no, no!
CHI7: Neck, neck (touches her neck). Esto [This]
CHI8: No
CHI7: Sí [Yes], neck
CHI8: I think is ‘sore throat’. Dolor de garganta [Sore throat]
CHI7: (Shakes her head)

	CHI8: No? CHI7: Neck is… CHI8: Bueno… [Well…] And his…	
	L-LRE (VP)	Pair 16, TG, Stage 4, Cycle 2
	CHI32: And he… 'salió' ['off']? 'salió volando' [flew off]. Out flying. CHI31: No, no, no (takes the pen and writes 'and the head was flying'). CHI32: Out flying! CHI31: Que no! [No!]	
	F-LRE (verb form)	Pair 4, CG, Stage 2, Cycle 1
	CHI7: '…One day morning at six o'clock Lucy…' (the student is revising the text) 'Was' o 'were'? CHI8: Were sleeping CHI7: No, was, was CHI8: No CHI7: Sí [yes], because is singular, only one CHI8: Vale [ok]	
	F-LRE (verb tense)	Pair 21, LTG, Stage 4, Cycle 2
Form	CHI41: Later, he mix… mixes the potion and drink CHI42: Drank CHI41: No, is present. Drink CHI42: Ah. Point (meaning 'period'). Suddenly…	
	F-LRE (subject)	Pair 26, LTG, Stage 1, Cycle 1
	CHI52: No, and Anna she's sleeping CHI51: 'She is' no. O [Either] 'she': 'She is', o… or 'Anna', no? CHI52: A ver [let's see], it's Monday morning and Anna… she is sleep…, no CHI51: No. You have… 'escribir'…? ['write'…?] CHI52: Write	

CHI51: Write ‘Anna’ or ‘she’
CHI52: O sea [That is], and Anna…
CHI51: ‘Anna’ and ‘she’ is the same
CHI52: Ah, vale [Ah, ok]. And Anna sleep

F-LRE (object) Pair 18, TG, Stage 3, Cycle 2

CHI36: Comma, and drink…
CHI35: No
CHI36: And drink…
CHI35: Drink them
CHI36: Then, drink…
CHI35: No, drink them, se bebe eso [he drinks it]

F-LRE (plural) Pair 5, CG, Stage 1, Cycle 1

CHI9: Brush her teeth
CHI10: No!
CHI9: Bueno [Well]
CHI10: Her teeth and her…
CHI9: No, teeth. Her teeth…
CHI10: And her…
CHI9: No, tooth
CHI10: Her… teeth
CHI9: No, no…
CHI10: Tooth-teeth!
CHI9: Pues eso [That’s what I’m saying]. Teeth.
CHI10: Her teeth (they laugh). And her… (touches her hair)
CHI9: Tooth is a… eh…
CHI10: Tooth
CHI9: In plural
CHI10: No

CHI9: Is… CHI10: It's tooth CHI9: Ay… Her… Tooth… Teeth (whispering, thinking) CHI10: Teeth CHI9: Teeth. And… Her… CHI10: Hair	
F-LRE (possessive marker)	Pair 6, CG, Stage 2, Cycle 1
CHI12: On the Maria's foot CHI11: On the foot of Maria… bueno… [well…] CHI12: No, Maria's foot CHI11: Vale, pues put 'Maria's' [Ok then, write 'Maria's']	
F-LRE (possessive pronoun)	Pair 13, TG, Stage 1, Cycle 2
CHI25: Strange… pass… in… in el… [the] CHI26: In her CHI25: Eso [That's it] CHI26: In her body CHI25: Sí [Yes] CHI26: His CHI25: Sí [Yes]. Something strange pass in… CHI26: In her… his body	
F-LRE (comparative or superlative)	Pair 6, CG, Stage 3, Cycle 2
CHI12: Ay, espera [wait] CHI11: Transform CHI12: Transform CHI11: In a strange cat CHI12: Stranger cat CHI11: Strange cat	

CHI12: Ok

F-LRE (pronoun) Pair 2, CG, Stage 4, Cycle 2

CHI4: And she drink…
CHI3: He drink (emphasis on the pronoun). Is one boy

F-LRE (adverb) Pair 17, TG, Stage 1, Cycle 2

CHI33: We have to put this picture, because it's three-four-five, not three-five… ehh… a ver…[let me see…] And finally… boom! Have a little explosion in her body
CHI34: Ok. But 'finally' here no, because we have to…
CHI33: Ah, `finally', no… ya [true], and finally… boom, have a little… 'and then'?
CHI34: Yes

F-LRE (word order) Pair 14, TG, Stage 1, Cycle 2

CHI27: And the title, the title?
CHI28: 'Josu and Peter'
CHI27: No, the…
CHI28: 'The magic potion'
CHI27: No, 'The potion fail'
CHI28: 'The fail potion'
CHI27: Yes

F-LRE (article) Pair 25, LTG, Stage 4, Cycle 1

CHI50: The girl is sleeping
CHI49: Yes, the girl… one girl, one girl
CHI50: In the first picture one girl…
CHI49: Sí [yes], one… One…
CHI50: Noooo (laughs)
CHI49: Sí [yes], one girl. Uno [One]
CHI50: One girl… (writing)

	F-LRE (preposition)	Pair 14, TG, Stage 1, Cycle 2
	CHI27: No, change… CHI28: 'On a cat' or 'in a cat'? CHI27: Eh… on, in… CHI28: In… to… CHI27: To a cat CHI28: No! 'In' or 'on'. On a cat… in a cat CHI27: No, 'on' CHI28: 'On' is (puts her hand on the table) 'on' CHI27: 'In' is 'in' (they laugh). Change… CHI28: Bueno [Well], on a, on a, on a dog CHI27: In a… in a cat! CHI28: In a cat (reads the whole sentence:) '…then his body start change in a cat'… On a cat CHI27: In a cat. In a cat. I think CHI28: No. Then her body start change on a cat CHI27: In a cat CHI28: I think is better 'on', but…	
Discourse	**D-LRE (coherence and cohesion)**	Pair 17, TG, Stage 3, Cycle 1
	CHI33: And continue sleeping CHI34: We can put 'but Sarah continue sleeping' CHI33: And Sarah, no? CHI34: With 'but', because it is something bad CHI33: Vale (Ok).	
	D-LRE (paragraphs)	Pair 24, LTG, Stage 4, Cycle 2
	CHI48: Another paragraph or the same? CHI47: I think the same CHI48: Ok. And…	

	CHI47: Mmm… Then…	
	D-LRE (stylistics)	Pair 26, Stage 4, Cycle 1
	CHI52: ‘Carlota’, because ‘she’ we put many times, no? CHI51: Ok, perfect. Carlota…	
Other	**Other**	Pair 14, TG, Stage 2, Cycle 1
	CHI28: Here (referring to the model) is more explain it and here (their text) no CHI27: Yes. And here (model) is very explicane it… CHI28: (Laughs). Explain! CHI27: Explain it and here (text) no CHI28: This is I´m saying. Yes… CHI27: No… Bueno… [Well…] CHI28: In this (model) the vocabulary is more… CHI27: More… CHI28: Is great than this CHI27: Yes.	

Once the episodes were identified and classified in the protocols, the following step was to engage in a combined analysis of all four stages which would allow us to trace the children's noticing processes across them. Although a thorough description of the features identified by each dyad in each one of the stages was not one of the aims of the present dissertation, the researcher attempted to provide a broad picture of the traceability of the children's noticing.

Accordingly, at Stage 1 (writing stage), each PFN (i.e., each episode corresponding to Stage 1) was coded for (i) the aspects attended to, (ii) the resolution of the episode (resolved: correctly/incorrectly or unresolved: addressed/ignored), and (iii) whether or not the episode was solvable, that is, if a plausible solution could be found in the model, or unsolvable, if there was no solution in the model.

(i) The resolution of the episode

The procedures that the children engaged in when trying to solve the problems they met while writing their drafts were also coded. An episode started when a participant raised a concern about language and was considered complete when the children either found a solution to the problem at issue or left it unresolved and continued with the writing process. Based on previous work (Basterrechea & Leeser, 2019; Calzada & García Mayo, 2020b; Collins & White, 2019; García Mayo & Azkarai, 2016), the outcome of episodes was coded on the basis of 'resolved' and 'unresolved', even if the solutions were not written down in the children's texts. The 'resolved' episodes were further classified into 'correctly' or 'incorrectly', while the 'unresolved' episodes were coded as 'addressed', if the learners discussed the problem but did not arrive at an outcome, or 'ignored', if the learner paid no attention to their partner and moved on with the writing process. This decision was made since we thought that a more detailed categorization might offer a deeper insight into the data. The following examples illustrate them.

(5) **Resolved correctly** L-LRE-verb. Pair 10, TG, Stage 1, Cycle 1

CHI20: Ah yes… (speaks unintelligibly). How do you say 'sonar'? ['ring'?]
CHI19: *Ring*

(6) **Resolved incorrectly** F-LRE-verb form (agreement). Pair 14, TG, Stage 3, Cycle 1

CHI27: And she stops

CHI28: Stop*s*? (emphasis on the 's')

CHI27: *Stop*

CHI28: *Ah, ok*

(7) **Unresolved addressed** L-LRE-pronoun. Pair 16, TG, Stage 2, Cycle 1

CHI32: Si digo… How do you say 'se'? [If I say… how do you say 'se' (Spanish reflexive pronoun)?]

CHI31: 'Drive'? No. 'Drive'?

CHI32: ¿Cómo? [How?]

CHI31: 'Drive'…

CHI32: ¿Cómo? ¿Así? (She writes it down). Esto es 'conducir' [How? Like this? This is 'drive']

CHI31: *No sé* [I don't know]

CHI32: *Bueno.* Martine… [Well, Martine…]

CHI31: Martine… How do you say 'escalera'? [Martine… How do you say 'stairs'?]

(8) **Unresolved ignored** L-LRE-noun. Pair 11, TG, Stage 1, Cycle 1

CHI21: But the clock… '*cosquillas"*? ['tickling'?] (gesticulates). But the clock…

CHI22: *Yes, yes. And she… and she… 'dormir"*? And she sleep. Before… o sea, After. [Yes, yes. And she… and she… 'sleep'? And she sleep. Before… I mean, after]

(ii) Solvability:

Having coded the resolution of the episodes, we also described whether or not the model provided a solution to the problematic features that the children had noticed in their written output. Only the transcripts of the treatment groups were coded for this aspect, since the self-correction group did not receive any treatment. An example is given below.

(9) **Solvable with the model** F-LRE-plural. Pair 14, TG, Stage 1, Cycle 1

CHI27: Her feet

CHI28: Foot… feet

CHI27: Foot

CHI28: Ok, feet are two, and foot one, or two? (Both laugh). Lorea! Start ringing… Sarah stop the clock

CHI27: With her… no, his, que me lío [I´m getting mixed up]

CHI28: With his feet, bueno [well], foot… no, feet

CHI27: Feet, es que… [Feet, the thing is…]

CHI28: Feet es una, foot dos [Feet is one, foot two]

The word 'foot' appears in the model provided at Stage 2, so this pair could potentially find a solution to this problem in the model text.

(10) **Not solvable with the model** M-LRE-spelling. Pair 13, TG, Stage 1, Cycle 1

CHI25: 'And prepairs to go to school' (while proofreading their text). Prepares to go to school. 'Prepares' sería [It's 'prepares']

CHI26: ¿No se escribe así? [is it not written like that?]

CHI25: Creo que no [I don't think so]

CHI26: ¿Y cómo se escribe? [And how do you write that?]

CHI25: No lo sé pero… [I don't know but…]

The word ‘prepares’ is not present in the model. Therefore, the children cannot solve their spelling problem through the model.

At Stage 2 (comparison stage), each FN (i.e., each episode corresponding to Stage 2) generated by the participants in the three groups was coded for (i) aspects attended to and (ii) whether it was a PFN or not. That is, we examined whether these features had been noticed at Stage 1. Table 4 provides an example. In addition, as the dyads in the CG corrected their own texts, we also described the resolution of their corrections.

Table 4. Noticing stages

Example	Was it a PFN?	FN
CHI24: And Carlota CHI23: No, and... (Touches his hair). And… And she eh… CHI24: Clean her… CHI23: Clean her hair CHI24: Her hair (CHI23 erases 'hair' and writes it correctly). With her hair. CHI23: With her… hair, vale [Ok]. How do you say this? (points at the second last vignette, where the girl is combing her hair) CHI24: Her hair and… I don't know	Yes	CHI24: 'Combs' (model) CHI23: Yes. We didn't put 'combs her hair' (model) CHI34: Yes. We didn't know the word
CHI38: We don't put the mechanic… the mechanic hand CHI37: Yes	No	-
CHI27: When the clock… CHI28: Eh… CHI27: But 'clock' is… CHI28: When the… CHI27: 'Clock' es de pared ['Clock' is for a wall], what… CHI28: No… CHI27: Ah, sí [Ah, yes], clock, clock CHI28: Clock start ringing. …in his bed when… his… 'Alarm clock'!, is 'alarm clock' CHI27: No CHI28: Yes! CHI27: Suena la alarma en el reloj? [Does the alarm ring on the clock?] CHI28: Yes! CHI27: No CHI28: It's a clock… with alarm CHI27: Pues eso [That's what I´m saying] CHI28: When his alarm clock… CHI27: When his clock… CHI28: Clock (starts writing) CHI27: Alarm CHI28: No! (both laugh) CHI27: Clock… CHI28: When his alarm clock… when his clock CHI27: Clock… ring	Yes	CHI27: Yes. And… CHI28: Look! Alarm-clock! CHI27: Ah síiii! [It's true!] CHI28: I told you! CHI27: And I told you the six a.m. CHI28: Yes CHI27: Ah! (mocking her)

Note: PFN=Problematic feature noticed

At Stage 3 (rewriting stage) and 4 (delayed post-test), each CRE and LRE identified in the data was coded for (i) the aspects attended to, (ii) the resolution of the episode and (iii) whether or not the feature was an incorporation (either from the model or the self-correction). The incorporated features were further classified into features noticed and incorporated (FNI), if they had been previously noticed or features unnoticed and incorporated (FUI), if the children incorporated a feature from the model which had not been explicitly mentioned.

Also, we examined whether the FNIs were traceable to Stage 1, Stage 2 or from both, that is, if these incorporations corresponded to PFNs, FNs or both. Table 5 illustrates this issue: The row named 'Writing 1 (Stage 1)' corresponds to what the children actually wrote in their first drafts. 'PFN' is what they noticed at Stage 1, 'FN' what they noticed at Stage 2, and 'Writing 2' and 'Writing 3' what they wrote at Stages 3 and 4, respectively. The column 'Incorporation Stage 1' refers to those incorporations coming exclusively from Stage 1, the column 'Incorporation Stage 2' to those coming from Stage 2, etc. So, if we take the first example, we can observe that one of the pairs had a problem with spelling at Stage 1 (PFN) and thus we classified the incorporation found in 'Writing 2' as originated at Stage 1, whereas, in the second example the feature incorporated in this second draft surfaced at Stage 2 (as a FN).

Table 5. Incorporations

	FNI			FUI
	Incorporation Stage 1	Incorporation Stage 2	Incorporation Stages 1 & 2	
Writing 1 (Stage 1)	One day a *cientific* made a potion	Her clock takes out a mechanic hand and starts touching her feet.	She washes her teeth and she *brushes her hair*.	Then, she go to the bathroom and *washes her teeth*.
PFN	CHI57: How do you write 'scientific'? CHI58: c-i-e-n-t-i-f-i-c CHI57: Nooo CHI58: Yes!! CHI57: Vale [Ok]	-	CHI25: Vale [Ok]. Later she goes to the bath CHI26: And… CHI25: Wash her… her stooth and brush the hair, no? CHI26: I don't know CHI25: Yes CHI26: Brushes?. CHI27: Yes	-
FN	-	CHI47: 'Feather'! Look! CHI48: Feather? CHI47: Feather! (laughs).	CHI26: And… espera [wait]… 'she brushes her teeth and combs her hair' (reading the model) and we put 'later she goes to the bath and brushes her hair…' (reading their draft) CHI25: And is 'comb'	-
Writing 2 (Stage 3)	One day, a *cientific* was in his laboratory	And the alarm take out a mechanic hand with a *feather* and starts touching her feet.	She brushes her teeth, she *combs her hair*	She go to the bathroom, *brush her teeth* and tidy up her hair.
Writing 3 (Stage 4)	-	-	She goes to the bathroom *to comb her hair*	Lucy go to the bathroom and *brushes her theeth*.

Note: FNI=Feature noticed and incorporated; FUI=Feature unnoticed and incorporated

Once the transcripts were coded accordingly, we proceeded with the analysis of the data. The total number of turns and episodes in each task were tallied. A turn began when a learner started talking and finished when his/her partner began a new utterance. Following García Mayo and Azkarai (2016), in order to make the numbers comparable under the same conditions, the incidence of LREs and CREs was analyzed considering proportions of the total number of turns in each episode to the total number of turns in each task.

After all episodes were identified and classified on the basis of their frequency, nature and resolution, the data were submitted to statistical analysis. Given that a larger sample size would allow us to obtain higher statistical power and decrease the margin of error, the six main categories (CREs, M-LREs, L-LREs, F-LREs, D-LREs and Other) were subjected to statistical analysis, but not the subcategories (adjectives, noun, stylistics, spelling, etc.) since the numbers were very small. Before the tests were run, normal distribution of the variables was checked and some of the data failed this assumption because of the low sample sizes, however, parametric statistics was used on the grounds that (i) for such complex factorial designs, even if assumptions are violated, nonparametric tests fall short (Norman, 2010) and (ii) according to Wilcox (2001), Norman (2010) and Larson-Hall (2010), the results obtained with parametric tests, more specifically with ANOVA (the statistical test chosen for this study) are robust for highly skewed non-normal distributions and sample sizes and are therefore likely to be accurate.

Accordingly, for research questions 1 and 2, two-sample binomial tests were run to look for between-group differences in the proportion of episodes and incorporations, and one-sample binomial tests were used for within-group differences in the proportion of episodes and incorporations as well as their resolution across cycles. To find out whether there is any change in the types of categories noticed (RQ1) and incorporated (RQ2) within stages and across cycles (time 1 vs. time 2), episodes were analyzed using a mixed ANOVA with a between-groups variable (group) and a within-groups variable (time). The post-hoc tests used to ascertain where the differences lay were Bonferroni, adjusting the alpha level to the number of comparisons to avoid the risk of Type 1 error. For related variables, the assumption of sphericity was verified with the Mauchly test

and in the case of not fulfilling the assumption, the Greenhouse-Geisser test was used to correct for lack of sphericity. The significance level was set at $\alpha = 0.05$.

Effect sizes were also calculated for each statistical procedure. In order to measure the effect size of one-way and mixed ANOVA, we used partial eta squared. Cohen (1988) suggested that from 0.06 to 0.1 be considered a small effect size, 0.15 represents a medium effect size and from 0.15 to 1 a large effect size. For post-hoc tests, we used the *d* family of effect sizes. According to Cohen (1992), the effect size of $d = 0.2$ can be considered a 'small' effect size, 0.5 represents a 'medium' effect size and 0.8 a 'large' effect size.

5.6.2 Analysis of written production

The participants' collaborative texts (180 texts in total) produced throughout the stages in both cycles were transcribed and analyzed following three procedures: Type of clause, CAF and holistic measures. To examine the impact of models on the children's written production in the short-term, the quality of the third draft (Stage 4) in relation to the first one (Stage 1) was measured in both cycles (draft 1 vs. draft 3 and draft 4 vs. draft 6). In order to examine the effect of the feedback in the long run, the first and the last draft (draft 1 vs. draft 6) were compared. Each one of the CAF measures along with the type of clause and global analyses were analyzed quantitively by running mixed ANOVAs to track the development of each group across the different drafts and to look for differences between groups.

5.6.2.1 Type of clause

Following Torras (2005) and Cánovas Guirao (2017), the participants' texts were divided into clausal units according to their degree of grammaticality. Three units were identified: pre-clause, proto-clause and clause. They are defined and exemplified as follows:

- **Pre-clause**: grammatically incorrect unit of language consisting of fragmented or distorted strings of words, at times incomplete, in which the meaning intention is not always apparent.
  - *And tought with the plum*
  - *And Miki now pick upp*

- **Proto-clause**: Linguistic unit in which the children's meaning intention is clear but which contains grammatical inaccuracies or gaps in the clausal unit.
  - *Then go to the family of Helen*
  - *Is time that go to school*

- **Clause**: grammatically accurate unit of language which may present a slight inaccuracy in spelling, lexis, grammar or concordance.
  - *And she go to the bathroom*
  - *One day morning at six o'clock, Lucy was sliping*

In order to track the children's writing development, the total number of units for each clause type was tallied and compared in their first and revised drafts in both cycles and across groups.

5.6.2.2 CAF measures

With the purpose of identifying potential progress in the linguistic acceptability and comprehensibility of the learners' written texts from their original to their revised texts in both cycles, we have based our analysis on the measures used by Torras (2005), Torras, Navés, Celaya and Pérez-Vidal (2006), Cánovas Guirao (2017) and Coyle and Roca de Larios' (2014), since they also analyzed the written output of child EFL learners. Therefore, these measures were classified into four main areas: (i) accuracy, (ii) fluency, (iii) grammatical complexity and (iv) lexical diversity.

(i) Accuracy

The accuracy of the children's texts was analyzed according to the procedure carried out by Cánovas Guirao's (2017). Accordingly, an error ratio was used to measure overall accuracy: [number of linguistic errors/total number of words] × 10. With Cánovas Guirao (2017), we decided to use a 10-word ratio rather than the 100-word ratio as the children's texts were relatively short (i.e., less than 100 words). Error ratios were computed and compared as displayed in Table 6 below.

Table 6. Example of error ratios (adapted from Cánovas Guirao, 2017)

ERROR RATIOS	
Original text	Revised text
Today it is Monday and the sun is going up Ana is sliping at 6.00 a.m. the clock has started to ring. Ana don't want to woke up but she has to go to the shool but her clock do things to Ana wake up. Now, Ana is brushing her head and washing her teets. Sudently, she takes her shool bag and goes to the shool.	Today is Monday morning and Martine is sleeping in her bed at six o´clock her clock´s alarm starts ringing but Martine isn´t want to wake up. Martine puts her feet on her pillow, but is sleep. Now, her clock takes a feather and starts touching her feet for she get up to go to school. Now Martine´s washing her teeth and combing her hair. She takes her school bag and goes to school.
Pair 29, LTG, Stage 1, Cycle 1	Pair 29, LTG, Stage 3, Cycle 1
Nº words: 66 Nº errors: 20 Error ratio: (20/66)x10=**3,03**	Nº words: 73 Nº errors: 8 Error ratio: (8/73)x10=**1,10**
Today (1) ~~it~~ is Monday and the sun is going up (2). Ana is (3) sl~~ie~~eping (4) and at 6.00 a.m. the clock (5) ~~has started~~ starts to ring. Ana (6) ~~don't~~ doesn't want to (7) w~~o~~ake up but she has to go to (8) ~~the~~ (9) ~~shool~~ school (10). (11) ~~but~~ her clock (12) do~~e~~s things to (13) ~~Ana~~ wake up Ana. Now, Ana is (14) ~~brushing~~ combing her (15) ~~head~~ hair and ~~washing~~ brushing her (16) ~~teets~~ teeth. (17) ~~Sudently~~ suddenly, she takes her (18) ~~shool~~ school bag and goes to (19) ~~the~~ (20) ~~shool~~ school.	Today is Monday morning and Martine is sleeping in her bed (1). at six o´clock her (2) ~~clock´s~~ alarm clock starts ringing but Martine (3) ~~isn´t~~ doesn't want to wake up. Martine puts her (4) ~~feet~~ foot on her pillow, but (5) ~~is sleep~~ continues sleeping. Now, her clock takes a feather and starts touching her (6) ~~feet~~ foot (7) ~~for she get up to go to school~~ to wake her up (because she has to go to school). Now Martine´s (8) ~~washing~~ brushing her teeth and combing her hair. She takes her school bag and goes to school.

(ii) Fluency

In order to give an idea of the children's fluency in writing, following Torras et al. (2006), the total number of words per text was also considered. An example is provided in Table 7. An online text analysis tool (https://textinspector.com/workflow) was used for the calculation of words.

(iii) Grammatical complexity

In line with previous research (Torras et al., 2006), grammatical complexity was measured as number of subordinate clauses and coordinate clauses. Examples of these categories are provided in Table 7.

Table 7. Example of fluency and grammatical complexity codification

One day a scientist was doing a potion while is dog was sleeping. When he finishes his potion, he was excited to taste it and he drinks the potion. When he drinks the potion he started feeling bad. Suddenly a bright light apier and a loud sound sounds and he turn into a cat. When the dog heard the sound it woke up and started fighting with the cat. 'Miau' were the last words of the scientist cat.	
Pair 13, TG, Stage 3, Cycle 2	
Subordinate clauses:	**Coordinate clauses:**
- *While his dog was sleeping* - *When he finishes his potion* - *When he drinks the potion* - *When the dog heard the sound*	- *and he drinks the potion.* - *and a loud sound sounds* - *and he turn into a cat.* - *and started fighting with the cat.*
Total word count: 78 Number of sentences: 6 Number of clauses: 14 Number of subordinate clauses: 4 Number of coordinate clauses: 4 Clauses per sentence: [14/6]=**2,33**	

(iv) Lexical diversity

Some extensively used measures such as lexical complexity (number of verbs, adjectives, noun, etc. types) as used by Torras et al. (2009) or lexical density (what proportion of the text contains lexical words) were discarded. This methodological decision was motivated by the fact that the focus of the present analysis was not to obtain the total number of lexical words or the different types of lexical categories

present in the children's texts, but rather to explore how many different words appear in each text. Consequently, lexical diversity (or lexical richness) was used as a measurement for newly learned vocabulary, since we considered that it might better reflect the development of the children's IL in terms of incorporation of new words.

Lexical diversity is usually calculated using a type-token ratio (TTR). This takes the number of 'types' (different words) and divides them by the 'tokens' (total number of words in the text). A high TTR indicates a high degree of lexical variation while a low TTR indicates the opposite. Although the TTR can be an extremely useful measurement for calculating the lexical diversity of a text, a common problem with this measure is that it does not work as effectively when dealing with texts of different length. However, Guiraud (1960) proposed a measure called 'Root Type Token Ratio' (RTTR) which is obtained by dividing the number of types by the square root of the number of tokens thus partially addressing the problem of TTR's variance on text length.

That said, each transcribed text was uploaded to a software tool (https://textinspector.com/workflow) which would calculate the number of types and tokens for us. Prior to uploading the children's written texts, spelling mistakes were corrected as the aim of lexical diversity is to analyze how diverse the range of words used is and therefore the software must recognize the words. Once this information was provided, we applied Guiraud's (1960) formula to the data (types/√tokens), thus obtaining the RTTR for each text.
The small size of most of the dataset meant that it was not possible to investigate the significant relationships between feedback treatments in terms of the potential development in the children's written output, so the data obtained was analyzed qualitatively.

5.6.2.3 Holistic measures

Finally, holistic measures were used as a complement to the quantitative analysis of the learners' written production, so as to give a wider view of the progress of the children's written output over an extended period of time. To this purpose, we assessed the texts both quantitatively and qualitatively taking into account measures of adequacy,

coherence, cohesion, grammatical accuracy, lexical range and mechanics. A three-point scoring rubric was used to evaluate the writings, 3 being good, 2 average and 1 poor (see Appendix 26). We made use of Villarreal and Munarriz-Ibarrola's (2021) rubric which is in turn an adaptation of the one used at the school their participants attended to as well as of the writing scale designed by Hedgcock and Lefkowitz (1992).

To ensure rating reliability, the participants' written production was coded by the researcher and 20% of the data (36 texts out of 180) was independently coded by the thesis supervisor. Inter-rater agreement resulted in 97%, and the remining differences were discussed until total agreement was reached.

5.6.2.4 Analysis of input enhancement

With the aim to explore the extent to which the enhanced linguistic features facilitated development in the knowledge of possessive pronouns, the total number of possessive pronouns used in each text was calculated and coded (correct/incorrect). One-sample and two-sample binomial tests were conducted to compare proportions and to examine the effects of input enhancement within and across cycles.

5.6.3 Analysis of the children's attitudes

Finally, in order to address the fifth research question, a mixed-method research design combining quantitative and qualitative data was used for the motivation thermometer as well as the attitude questionnaires and interviews. Quantitative data were obtained from the motivation thermometers' responses of each student before and after the task (i.e., the Likert scale questions alongside the reasons provided) as well as from the close-ended questions in the anonymous questionnaires. This served us to examine the children's initial and final motivation towards the different feedback conditions. As usual, a mixed ANOVA was performed, since we are combining features of both a between-subjects design and a within-subjects design.

Qualitative data were obtained from the open-ended questions in the survey and the information gathered from the interviews. Responses were compared, and recurrent patterns or themes identified across the different interviews with the focus groups.

CHAPTER 6: RESULTS

In this chapter the results corresponding to the five research questions posited in Chapter 5 will be presented. As shown below, and for a clearer understanding, the research questions have been organized by modules in such a way that (i) the first module comprises the first two questions, which deal with the oral-written connection; (ii) the second module encompasses the third and fourth questions, which are concerned with the effects of models on the children's writing; and (iii) the third module embraces the question on learners' attitudes and motivation. The evidence obtained has been analyzed mainly from a quantitative perspective, except for the answer to research question 5, for which we have followed a mixed-method design. In order to make this section more reader friendly, only significant results will be highlighted, although non-significant results will also be provided in the tables (statistical significance has been marked with an asterisk).

6.1. Module 1: Oral-written connection

This first module revolves around the impact that child oral interaction has on their written product. Accordingly, in what follows, we will present the results obtained for the first research question, for which we conducted a quantitative analysis of the episodes reported on the children's noticing as well as of their traceability across stages and into the revised texts.

6.1.1 Results for research question 1

The first research question is stated here again for the reader's convenience:

1. What linguistic features do young EFL children focus on when dealing with a task using model texts?

a) What features of language do they notice when composing a narrative text?
b) What do they notice when comparing their written texts with model texts or when correcting themselves?

c) Are there any across-cycle differences within groups (CG, TG and long-term treatment group (LTG)) regarding frequency, type and outcome of the episodes produced in oral interaction?

As shown above, the first set of questions encompasses three main aspects. Research question 1a aims to identify the features that the children noticed the most or found most problematic (PFNs) when writing a text in collaboration. Table 8 shows the frequencies, means and proportions of PFNs **at Stage 1 (composing stage) of Cycle 1**. A total of 154 episodes out of 513 (29.50%) were generated by the dyads in the CG, the TG produced 186 episodes (36.26%), while 173 (33.14%) corresponded to the pairs in the LTG, so the total number of features was rather equally distributed between groups. Actually, no statistically significant differences were found between groups for Stage 1, Cycle 1 (CG vs. TG: $z = 1.42$; $p = .152$. CG vs. LTG: $z = 1.27$; $p = .203$. TG vs. LTG: $z = 0.85$; $p = .394$). Regarding the nature of the PFNs, at first glance the results indicate that most learner pairs, irrespective of the feedback condition they were assigned to, encountered mainly formal (29.63%), lexical (29.04%) and mechanical (24.56%) problems at the moment of writing rather than content (10.92%), discourse (5.07%) or other issues (0.78%).

Table 8. Frequencies, Means and Proportions of PFNs at Stage 1, Cycle 1

	All participants (*N*=30)				**CG (*N*=9)**				**TG (*N*=11)**				**LTG (*N*=10)**			
	FO[1]	*%*	*Mean*	*SD*	*FO*	*%*	*Mean*	*SD*	*FO*	*%*	*Mean*	*SD*	*FO*	*%*	*Mean*	*SD*
CREs	56	10.92	1.87	1.43	12	7.79	1.33	1	27	14.52	2.45	1.57	17	9.83	1.7	1.49
Mechanics	126	24.56	4.2	3.28	42	27.27	4.67	3.04	37	19.89	3.36	2.94	47	27.17	4.7	3.95
Lexis	149	29.04	4.9	2.89	46	29.87	4.89	3.22	63	33.87	5.73	2.87	40	23.12	4	2.62
Form	152	29.63	5.07	4.10	46	29.87	5.11	4.48	49	26.34	4.45	4.06	57	32.95	5.7	4.14
Discourse	26	5.07	0.87	1.07	6	3.90	0.67	0.71	9	4.84	0.82	0.87	11	6.36	1.1	1.52
Other	4	0.78	0.13	0.43	2	1.30	0.22	0.67	1	0.54	0.09	0.30	1	0.58	0.1	0.32
Total	**513**	100	2.84	3.20	**154**	29.50	2.81	3.29	**186**	36.26	2.82	3.09	**173**	33.14	2.88	3.31

[1]Frequency of occurrences

Let us now zoom in on the statistical analyses for Stage 1. Sphericity was verified for Categories (X^2 = 84.29; p = < .000) and for the interaction between Categories and Cycles (X^2 = 65.89; p = < .000). Given that the sphericity assumption was not met for related variables, the Greenhouse-Geisser correction was used. Table 9 shows the results of the mixed ANOVA for Stage 1. As can be seen, mixed ANOVA revealed significant effects for Categories ($F_{(2.62,70.75)}$ = 33.39; p = < .000; η_p^2 = 0.55), Cycles ($F_{(1,27)}$ = 9.86; p = .004; η_p^2 = 0.26), Group ($F_{(2,27)}$ = 23.42; p = .003; η_p^2 = 0.33) and for the interaction between Categories and Group ($F_{(5.24,70.74)}$ = 7.32; p = .042; η_p^2 = 0.17), Categories and Cycles ($F_{(3.15,85.22)}$ = 4.45; p = .005; η^2 = 0.14) and Categories, Cycles and Group ($F_{(6.31,85.22)}$ = 5.34; p = < .000; η^2 = 0.10). Multiple comparisons located these differences across categories for each group in both Cycles 1 and 2, and between groups only in Cycle 2.

Table 9. Mixed-model ANOVA results for Stage 1

Stage 1					
	***df*1**	***df*2**	***F***	***p***	η_p^2
Categories	2.62	70.75	33.39	< .000*	0.55
Categories * Group	5.24	70.74	7.32	.042*	0.17
Cycles	1	27	9.86	.004*	0.26
Cycles * Group	2	27	1.44	.255	0.09
Categories * Cycles	3.15	85.22	4.45	.005*	0.14
Categories * Cycles * Group	6.31	85.22	5.34	<.000*	0.10
Inter-subjects					
Group	2	27	23.42	.003*	0.33

As for Cycle 1, the results confirm the information stated above. That is, while no differences were observed between content and discourse and discourse and other issues in any of the three groups, lexis, form and mechanics did receive significantly more attention than the rest of the categories. Statistically significant differences are also observable between content and other, given the reduced number of items contained in the latter, and there are also medium and large effect sizes for the difference between discourse and other in the three groups, which means that this difference can be considered important. Table 10 details these findings:

Table 10. Results for within-group differences across categories at Stage 1, Cycle 1

Group	Categories	*p*	*d*	*95% Confidence Interval*	
				Lower bound	**Upper bound**
CG	Content-Mechanics	.037*	-1.47	-6.55	-0.12
	Content-Lexis	.005*	-1.49	-6.33	-0.78
	Content-Form	.036*	-1.16	-8	0.45
	Content-Discourse	1.00	0.57	-0.84	2.17
	Content-Other	.017*	1.31	-0.35	2.57
	Mechanics-Lexis	1.00	-0.07	-3.25	2.8
	Mechanics-Form	1.00	-0.12	-4.81	3.92
	Mechanics-Discourse	.014*	1.81	0.54	7.46
	Mechanics-Other	.005*	2.02	0.96	7.93
	Lexis-Form	1.00	-0.06	-4.97	4.52
	Lexis-Discourse	.008*	1.81	0.78	7.66
	Lexis-Other	.001*	2.01	1.53	7.81
	Form-Discourse	.038*	1.38	0.15	8.74
	Form-Other	.023*	1.52	0.42	9.36
	Discourse-Other	1.00	0.65	-0.8	1.69
TG	Content-Mechanics	.041*	-0.39	-3.82	2
	Content-Lexis	.004*	-1.42	-5.78	-0.76
	Content-Form	.044*	-0.65	-5.83	1.83
	Content-Discourse	.062	0.29	0.28	3
	Content-Other	<.000*	2.09	1.05	3.68
	Mechanics-Lexis	.146	-0.31	-6	0.37
	Mechanics-Form	1.00	-0.31	-5.04	2.86
	Mechanics-Discourse	.013*	1.17	-0.59	5.68
	Mechanics-Other	.037*	1.57	0.12	6.43
	Lexis-Form	1.00	0.36	-3.02	5.57
	Lexis-Discourse	<.000*	2.32	1.8	8.02
	Lexis-Other	<.000*	2.76	2.79	8.48
	Form-Discourse	.016*	1.24	-0.25	7.52
	Form-Other	.026*	0.02	0.32	8.4
	Discourse-Other	.711	1.11	-0.4	8.4
LTG	Content-Mechanics	.028*	-1.01	-6.05	0.05
	Content-Lexis	.014*	-1.08	-4.93	0.33
	Content-Form	.019*	-1.29	-8.01	0.01
	Content-Discourse	1.00	0.40	-0.83	2.03
	Content-Other	.014*	1.48	0.22	2.98
	Mechanics-Lexis	1.00	0.21	-2.17	3.57
	Mechanics-Form	1.00	-0.25	-5.14	3.14
	Mechanics-Discourse	.023*	1.20	0.32	6.89
	Mechanics-Other	.002*	1.64	1.29	7.91
	Lexis-Form	1.00	-0.49	-6.2	2.8
	Lexis-Discourse	.016*	1.35	-0.36	6.16
	Lexis-Other	.004*	2.10	0.92	6.88
	Form-Discourse	.017*	1.48	0.52	8.68
	Form-Other	.003*	1.91	1.36	9.84
	Discourse-Other	.168	0.93	-0.18	2.18

The fact that 83% of the PFNs at Stage 1 were formal, lexical and mechanical leads one to believe that the children's attention at this first stage of the task was focused predominantly on finding the words to convey their intended meanings and on being accurate. More precisely, as far as lexis is concerned, the children experienced problems mostly with verbs and nouns; regarding form, verb form (agreement), verb tense and possessive pronouns were the most attended to, while in the case of mechanics, spelling and punctuation were the children's primary focus of attention (See Appendix 27 for a detailed table). In sum, we observe that, when composing a narrative text in collaboration, the child learners attended mostly to form, lexis and mechanics and that the groups were similar insofar as the focus of their attention, since no significant differences were observed between groups.

If we now turn to **Stage 1 of Cycle 2**, we would expect groups to behave differently as some children kept self-correcting their texts for months, some others benefitted from models during this same period, while the remaining pairs had only received models once by now. Table 11 shows the frequencies, means and proportions of PFNs at Stage 1 of Cycle 2. Out of a total of 360 episodes, the CG identified 135 (37.5%) PFNs, the TG verbalized 137 episodes (38.05%), whereas the learner pairs in the LTG encountered 88 PFNs (24.44%). This time, the binomial test showed significant differences between the LTG and the CG ($z = 3.79$, $p = < .000$) and between the LTG and the TG ($z = 3.94$, $p = < .000$), but not between the CG and the TG ($z = 0.15$, $p = .877$), which implies that the children in the LTG verbalized significantly fewer episodes than their counterparts. As for the types of episodes, consistent with Cycle 1 noticing, the largest proportion of overall PFNs were mechanical (30.55%), formal (28.33%) and lexical (24.17%) when writing their first draft. Let us now consider what each group did individually.

Table 11. Frequencies, Means and Proportions of PFNs at Stage 1, Cycle 2

	All participants (*N*=30)				**CG (*N*=9)**				**TG (*N*=11)**				**LTG (*N*=10)**			
	FO	*%*	*Mean*	*SD*	*FO*	*%*	*Mean*	*SD*	*FO*	*%*	*Mean*	*SD*	*FO*	*%*	*Mean*	*SD*
CREs	25	6.94	0.83	0.98	6	4.44	0.67	0.71	16	11.68	1.45	1.13	3	3.41	0.3	0.67
Mechanics	110	30.55	3.67	3.12	54	40	6	3.12	35	25.55	3.18	2.79	21	23.86	2.1	2.38
Lexis	87	24.17	2.9	2.14	28	20.74	3.11	1.76	37	27	3.36	2.80	22	25	2.2	1.55
Form	102	28.33	3.4	3.19	40	29.63	4.44	4.39	36	26.28	3.27	2.72	26	29.54	2.6	2.37
Discourse	28	7.78	0.93	0.98	3	2.22	0.33	0.5	11	8.03	1	0.77	14	15.91	1.4	1.26
Other	8	2.22	0.27	0.58	4	2.96	0.44	0.73	2	1.46	0.18	0.60	2	2.27	0.2	0.42
Total	**360**	100	2	2.49	**135**	37.5	2.5	3.15	**137**	38.05	2.07	2.34	**88**	24.44	1.47	1.82

As happened in Cycle 1, pairwise comparisons showed significant within-group differences across categories in Cycle 2. The analysis revealed similar results to those obtained in Cycle 1, namely, mechanics, lexis and form were significantly more attended to than the rest of the categories, at least for both the CG and the TG (although the effect size for the comparison between mechanics and lexis in the CG is very large and also shows a difference close to significance, which implies that there is a strong tendency for mechanics to be more attended to than lexis). As for the LTG, no differences were found between content and other, but discourse seemed to have gained strength and is now added to the equation along with mechanics, lexis and form as the most significantly noticed categories at the moment of writing in Cycle 2. A table for this data is shown below.

Table 12. Results for within-group differences across categories at Stage 1, Cycle 2

Group	Categories	*p*	*d*	*95% Confidence Interval*	
				Lower bound	**Upper bound**
CG	Content-Mechanics	<.000*	-2.36	-8.09	-2.58
	Content-Lexis	.010*	-1.82	-4.49	-0.4
	Content-Form	.027*	-1.20	-7.29	-0.27
	Content-Discourse	1.00	0.54	-0.97	1.63
	Content-Other	1.00	0.31	-0.62	1.07
	Mechanics-Lexis	.054	1.14	-0.03	5.81
	Mechanics-Form	1.00	0.41	-2.21	5.32
	Mechanics-Discourse	<.000*	2.53	2.52	8.81
	Mechanics-Other	<.000*	2.45	2.89	8.23
	Lexis-Form	1.00	-0.40	-4.94	2.28
	Lexis-Discourse	.016*	2.14	0.34	5.22
	Lexis-Other	.011*	1.98	0.41	4.93
	Form-Discourse	.009*	1.32	0.7	7.52
	Form-Other	.019*	1.27	0.42	7.59
	Discourse-Other	1.00	-0.18	-1.28	1.06
TG	Content-Mechanics	.037*	-0.81	-4.22	0.76
	Content-Lexis	.039*	-0.89	-3.76	-0.06
	Content-Form	.041*	-0.87	-5	1.36
	Content-Discourse	1.00	0.47	-0.72	1.63
	Content-Other	<.000*	1.41	0.51	2.04
	Mechanics-Lexis	1.00	-0.07	-2.82	2.46
	Mechanics-Form	1.00	-0.03	-3.49	3.31
	Mechanics-Discourse	.032*	1.07	-0.66	5.03
	Mechanics-Other	.007*	1.49	0.59	5.42
	Lexis-Form	1.00	0.03	-3.17	3.36
	Lexis-Discourse	.028*	1.15	0.16	4.57
	Lexis-Other	<.000*	1.57	1.14	5.23
	Form-Discourse	.029*	1.14	-0.81	5.36
	Form-Other	.001*	1.57	-0.15	6.33
	Discourse-Other	.044*	1.18	-0.24	1.88
LTG	Content-Mechanics	.047*	-1.03	-4.41	0.81
	Content-Lexis	.039*	-1.59	-3.84	0.04
	Content-Form	.012*	-1.32	-5.63	1.03
	Content-Discourse	.031*	-1.08	-2.33	0.13
	Content-Other	1.00	0.18	-0.7	0.9
	Mechanics-Lexis	1.00	-0.05	-2.87	2.67
	Mechanics-Form	1.00	-0.21	-4.07	3.07
	Mechanics-Discourse	1.00	0.37	-2.28	3.68
	Mechanics-Other	.034*	1.11	-0.63	4.43
	Lexis-Form	1.00	-0.20	-3.82	3.02
	Lexis-Discourse	1.00	0.57	-1.52	3.12
	Lexis-Other	.002*	1.76	-0.14	4.14
	Form-Discourse	1.00	0.63	-2.04	4.44
	Form-Other	.018*	1.41	-1	5.8
	Discourse-Other	.026*	1.27	0.09	2.31

Taken together, these results insinuate that in this first stage of the second cycle the children in the CG and the TG continued experiencing mostly mechanical, lexical and formal problems. Nevertheless, apart from these three categories, for some reason, the LTG also seems to allocate significantly more attentional resources to discourse as opposed to Cycle 1. In terms of which parts of speech the children were most concerned with (see Appendix 28), spelling and punctuation were again the primary focus of attention as far as mechanics is concerned; verbs and nouns appertaining to lexis, verb tense and possessive pronouns in the case of form and stylistic issues with regard to discourse.

In this case, we have between-group differences. Post-hoc comparisons located the differences shown by the mixed ANOVA between the TG and the LTG for content, and between the CG and the LTG for mechanics, as shown in Table 13. This finding seems to indicate that, after being subjected to different treatments, the participants in the three feedback conditions continued focusing on lexis, form, discourse and other issues to a similar extent during the composing stage. Nevertheless, the problems related to content clarification or idea generation that the TG encountered were found to be greater than those observed by the LTG. In like fashion, we found the self-correction group to be more interested in mechanics than the children in the LTG.

Table 13. Results for between-group differences at Stage 1, Cycle 2

Categories per group	**POST-HOC**				
	Group	***p***	***d***	***95% Confidence Interval***	
				Lower bound	**Upper bound**
CG-TG-LTG: Content	CG-TG	.168	-0.36	-1.80	0.22
	TG-LTG	.017*	0.55	0.18	2.13
	CG-LTG	1.00	0.17	-0.66	1.40
CG-TG-LTG: Mechanics	CG-TG	.095	0.41	-0.36	5.99
	TG-LTG	1.00	0.16	-2	4.17
	CG-LTG	.015*	-0.56	-7.14	-0.66

Overall, these results provide important insights into the children's focus of attention during the first stages before and after receiving different treatments. Both cycles happen to be characterized by a powerful noticing of grammar, vocabulary, spelling and punctuation problems when writing their texts in response to a visual prompt. In Cycle

1, the three groups behave similarly regarding the number of episodes produced and the types of features they find troublesome. Nonetheless, we find some discrepancies across categories and groups in Cycle 2 which lead us to think of some aftereffect of the feedback treatment received by each group. For example, the LTG verbalized considerably fewer episodes than their counterparts. In addition, this group also appeared to show concern about such matters as stylistics, paragraphs or coherence and cohesion (discourse) to a greater extent than in Cycle 1. We also observe differences across groups, as in Cycle 2 the TG attended more to CREs than the LTG, and the CG focused significantly more on mechanics than their classmates in the LTG. Notwithstanding this first approach, more information is needed to obtain and understand the full picture. In the following lines we will give answer to research question 1b, which aimed to identify the features that the children noticed the most (FNs) when comparing their drafts with the model and when correcting their own texts.

Following the same procedure and statistical methods as those used for the previous question, Table 14 presents the results obtained from the preliminary analysis of the FNs **at Stage 2 of Cycle 1**. The total number of FNs at Stage 2 (n = 226) was lower than at Stage 1 (n = 513). Nevertheless, we now find a large difference between the treatment groups and the self-correction group which was not present at Stage 1. On the one hand, the TG noticed 118 features (52.21%) and the LTG identified 108 (47.79%) while comparing their drafts to the model texts, with no significant differences between them ($z = 0.94$, $p = .346$). On the other hand, the learner dyads in the CG verbalized a lower amount of LREs and CREs (n = 51; 22.57%) than both the TG ($z = 6.51$, $p = < .000$) and the LTG ($z = 5.61$, $p = < .000$). With respect to categories, the treatment groups appeared to be more concerned with lexical and content-related features (60% for TG and 70% for LTG), whereas the CG attended mostly to issues related to form (43.14%) and, to a lesser extent, mechanics (23.53%).

Table 14. Frequencies, Means and Proportions of FNs at Stage 2, Cycle 1

	All participants (*N*=30)				**CG (*N*=9)**				**TG (*N*=11)**				**LTG (*N*=10)**			
	FO	*%*	*Mean*	*SD*	*FO*	*%*	*Mean*	*SD*	*FO*	*%*	*Mean*	*SD*	*FO*	*%*	*Mean*	*SD*
CREs	81	35.84	2.7	2.23	2	3.92	0.22	0.44	42	35.59	3.82	1.66	37	34.26	3.7	2
Mechanics	21	9.29	0.7	1.05	12	23.53	1.33	1.5	4	3.39	0.36	0.50	5	4.63	0.5	0.85
Lexis	81	35.84	2.7	2.64	6	11.76	0.67	0.71	37	31.35	3.36	3.11	38	35.19	3.8	2.30
Form	49	21.68	1.63	2.09	22	43.14	2.44	3.13	12	10.17	1.09	1.76	15	13.89	1.5	0.97
Discourse	29	12.83	0.97	0.85	9	17.65	1	1	12	10.17	1.09	0.83	8	7.41	0.8	0.79
Other	16	7.08	0.53	0.82	0	0	0	0	11	9.32	1	1	5	4.63	0.5	0.70
Total	**226**	100	1.54	1.96	**51**	22.57	0.94	1.65	**118**	52.21	1.79	2.10	**108**	47.79	1.8	1.97

In terms of statistical analyses, sphericity was verified for Categories (X^2 = 31.28; p = .005) and for Categories x Cycles interaction (X^2 = 32.31; p = .004). The assumption of sphericity was violated, so the Greenhouse-Geisser correction was used. Table 15 shows the results of the mixed ANOVA for Stage 2. As can be seen from the table below, statistical tests revealed that there is a main effect for Categories ($F_{(3.54,95.67)}$ = 13.36; p = < .000; η_p^2 = 0.33), Cycles ($F_{(1,27)}$ = 13.84; p = .001; η_p^2 = 0.33), Group ($F_{(2,27)}$ = 4.61; p = .019; η_p^2 = 0.25) and a significant interaction effect between Categories and Group ($F_{(7.08,95.66)}$ = 9.54; p = .019; η_p^2 = 0.25), Cycles and Group ($F_{(2,27)}$ = 16.25; p = < .000; η_p^2 = 0.55), Categories and Cycles ($F_{(3.43,92.52)}$ = 5.34; p = .001; η_p^2 = 0.17), and between Categories, Cycles and Group ($F_{(6.85,92.52)}$ = 5.03; p = < .000; η_p^2 = 0.27). Further post-hoc tests located these differences across categories for each group and between groups at Stage 2 of Cycles 1 and 2, as explained below.

Table 15. Mixed-model ANOVA results for Stage 2

Stage 2					
	***df*1**	***df*2**	***F***	***p***	η_p^2
Categories	3.54	95.67	13.36	< .000*	0.33
Categories * Group	7.08	95.66	9.54	.019*	0.25
Cycles	1	27	13.84	.001*	0.33
Cycles * Group	2	27	16.25	< .000*	0.55
Categories * Cycles	3.43	92.52	5.34	.001*	0.17
Categories * Cycles * Group	6.85	92.52	5.03	< .000*	0.27
Inter-subjects					
Group	2	27	4.61	.019*	0.25

From Table 16, which shows the post-hoc results for Cycle 1, we can see that the pairs in the CG focused significantly more on formal aspects of the language than on any other category when correcting their own texts. Mechanics is also one of the categories that the children paid most attention to (n = 12), but it is only significantly different from content, lexis and other. Interestingly, paired t-tests confirm the above-mentioned results for the groups that dealt with the model. In general terms, both lexis and content proved significantly different from the remaining categories. However, unlike the CG, form and mechanics is pushed into the background to a statistically significant degree. As a matter of fact, in the case of the TG, mechanics was found to be even significantly less frequent than discourse and other issues, and no differences between form and

discourse and form and other were found. In other words, the categories which normally receive the least attention (discourse and other) now surpass formal and mechanical aspects. A similar situation can be found with reference to the LTG, who focused primarily on meaning, thus consigning mechanics, form and discourse to oblivion.

Table 16. Results for within-group differences across categories at Stage 2, Cycle 1

Group	Categories	*p*	*d*	*95% Confidence Interval*	
				Lower bound	**Upper bound**
CG	Content-Mechanics	.007*	-1.62	-2.90	0.68
	Content-Lexis	.032*	-0.75	-3.42	2.53
	Content-Form	.008*	-2.59	-4.80	0.35
	Content-Discourse	.045*	-1.01	-2.79	1.23
	Content-Other	1.00	0.71	-1.56	2.01
	Mechanics-Lexis	.030*	0.84	-1.93	3.26
	Mechanics-Form	1.00	0	-2.99	1.00
	Mechanics-Discourse	1.00	0.36	-0.94	1.61
	Mechanics-Other	.004*	2.18	0.31	2.36
	Lexis-Form	.008*	-1.89	-4.44	0.88
	Lexis-Discourse	1.00	-0.38	-3.11	2.44
	Lexis-Other	.023*	1.33	-1.88	3.22
	Form-Discourse	.001*	1.35	-0.11	2.99
	Form-Other	<.000*	3.06	0.94	3.95
	Discourse-Other	.046*	1.41	0.01	1.99
TG	Content-Mechanics	<.000*	2.81	1.83	5.07
	Content-Lexis	1.00	0.18	-2.24	3.15
	Content-Form	.012*	1.59	0.40	5.06
	Content-Discourse	.001*	2.08	0.91	4.54
	Content-Other	<.000*	2.05	1.20	4.43
	Mechanics-Lexis	.005*	-1.35	-5.35	-0.65
	Mechanics-Form	1.00	-0.56	-2.43	1.00
	Mechanics-Discourse	.028*	-1.06	-1.88	0.43
	Mechanics-Other	.032*	-0.80	-1.56	0.29
	Lexis-Form	.039*	0.90	-0.13	4.68
	Lexis-Discourse	.047*	1.00	-0.23	4.78
	Lexis-Other	.031*	1.02	0.06	4.67
	Form-Discourse	1.00	0	-1.40	1.40
	Form-Other	1.00	0.06	-1.27	1.45
	Discourse-Other	1.00	0.10	-0.81	1.00
LTG	Content-Mechanics	<.000*	2.08	1.50	4.90
	Content-Lexis	1.00	-0.05	-2.92	2.72
	Content-Form	.021*	1.40	-0.24	4.64
	Content-Discourse	.001*	1.91	0.99	4.81
	Content-Other	<.000*	2.13	1.51	4.89
	Mechanics-Lexis	.003*	-2.14	-5.76	-0.84
	Mechanics-Form	1.00	-1.10	-2.79	1.22
	Mechanics-Discourse	1.00	-0.37	-1.51	1.98
	Mechanics-Other	1.00	0	-0.97	1.35
	Lexis-Form	.018*	1.30	-0.22	4.82
	Lexis-Discourse	.016*	1.74	0.37	5.63
	Lexis-Other	.002*	1.94	0.88	5.72
	Form-Discourse	1.00	0.29	-0.77	2.17
	Form-Other	.035*	1.18	-0.43	2.43
	Discourse-Other	1.00	0.40	-0.64	1.24

The types of features noticed by the children in the self-correction group were related mainly to spelling, to the correct formation of verb tenses as well as morphological features such as subject-verb agreement or even the coherence, cohesion and stylistics of the text. On the contrary, a closer look at the FNs reported by the dyads assigned to the model condition indicated a special attention to content clarification as well as to new vocabulary identified in the model, especially verbs and nouns (see Appendix 29).

Table 17 shows the post-hoc test results for the inter-group differences at Stage 2 of Cycle 1. Comparisons using the Bonferroni correction found statistically significant differences between the treatment groups and the CG regarding content, mechanics and lexis and between the TG and the CG concerning other issues. More specifically, the treatment groups significantly outperformed the CG in their noticing of content and lexis, while the self-correction group noticed significantly more mechanical aspects than their counterparts. Therefore, from these results it can be inferred that models had an apparent effect on the amount of CREs and lexical LREs produced, but also that self-correction had a significant impact on the noticing of formal and mechanical aspects.

Table 17. Results for between-group differences at Stage 2, Cycle 1

Categories per group	**POST-HOC**				
	Group	***p***	***d***	***95% Confidence Interval***	
				Lower bound	**Upper bound**
CG-TG-LTG: Content	CG-TG	<.000*	-0.94	-5.38	-1.81
	TG-LTG	1.00	0.03	-1.62	1.85
	CG-LTG	<.000*	-0.89	-5.30	-1.65
CG-TG-LTG: Mechanics	CG-TG	.023*	0.83	0.11	1.83
	TG-LTG	1.00	-0.08	-0.97	-0.70
	CG-LTG	.048*	0.44	-0.04	1.71
CG-TG-LTG: Lexis	CG-TG	.049*	-0.47	-5.38	-0.01
	TG-LTG	1.00	-0.08	-3.05	2.18
	CG-LTG	.021*	-0.53	-5.88	-0.39
CG-TG-LTG: Other	CG-TG	.016*	-0.55	-1.84	-0.16
	TG-LTG	.390	0.29	-0.32	1.32
	CG-LTG	.447	-0.27	-1.36	0.36

From the above we could conclude that (i) the treatment groups noticed a considerably higher number of features than the CG; (ii) the treatment groups noticed significantly more lexical and content-related features than any other category and the difference was

also significant with respect to the CG's noticing of the same features; and (iii) the CG focused mainly on form and mechanics, the difference of the latter also being significant with respect to the treatment groups' noticing of mechanics. Therefore, the model text seemed to prompt a higher number of lexical and content-related episodes, while self-correction triggered a predominant focus on form and mechanics.

Finally, let us now analyze what children focused on **at Stage 2** after receiving different treatments. It is apparent from Table 18 that the three groups are now different. To start with, although the total number of episodes produced by all pairs (n = 234) is similar to those generated in Cycle 1 (n = 226), we can observe a sharp drop in the number of episodes identified by the groups who had received their corresponding treatment for months, that is, the CG (n = 52) and the LTG (n = 71), while the TG appears to continue along the same line (n = 111). Statistics corroborates the difference between these two sets of groups ($z = 5.72$, $p = .00$ for the comparison between TG and CG, and $z = 3.79$, $p = < .000$ for the comparison between TG and LTG). Still, the LTG showed differences with respect to the CG ($z = 1.99$; $p = .046$).

Moving on to consider the nature of the episodes, we also find great disparity among these results. The CG kept their concern about mechanics (38.46%) and form (26.92%) while trying to correct their texts. For their part, the children in the TG attended mostly to CREs (26.13%) and lexis (24.32%) as happened in Cycle 1, but also to mechanics (22.52%). The dyads in the LTG, however, were mostly engaged with issues related to content (25.35%), form (22.53%), mechanics (18.31%) and lexis (18.31%).

Table 18. Frequencies, Means and Proportions of FNs at Stage 2, Cycle 2

	All participants (*N*=30)				**CG (*N*=9)**				**TG (*N*=11)**				**LTG (*N*=10)**			
	FO	*%*	*Mean*	*SD*	*FO*	*%*	*Mean*	*SD*	*FO*	*%*	*Mean*	*SD*	*FO*	*%*	*Mean*	*SD*
CREs	51	21.79	1.7	1.9	4	7.69	0.44	0.53	29	26.13	2.64	2.06	18	25.35	1.8	1.99
Mechanics	58	24.79	1.93	1.46	20	38.46	2.22	1.71	25	22.52	2.27	1.42	13	18.31	1.3	1.16
Lexis	45	19.23	1.5	1.43	5	9.61	0.55	0.88	27	24.32	2.45	1.57	13	18.31	1.3	1.06
Form	44	18.80	1.47	1.28	14	26.92	1.55	1.81	14	12.61	1.27	1.10	16	22.53	1.6	0.97
Discourse	15	6.41	0.5	0.82	6	11.54	0.67	1	6	5.40	0.54	0.82	3	4.22	0.3	0.67
Other	21	8.97	0.7	0.91	3	5.77	0.33	0.5	10	9.01	0.91	0.94	8	11.27	0.8	1.13
Total	**234**	100	1.29	1.38	**52**	22.22	0.94	1.16	**111**	47.43	1.68	1.56	**71**	30.34	1.18	1.28

Looking at each group's response, statistical tests showed significant differences across categories in Cycle 2, as shown in Table 19. In particular, post-hoc tests located a significant difference for the CG between mechanics and the rest of the categories (except for form), and form was also found to be significantly different from lexis, content and other issues, but not from discourse. In relation to the TG, the results were statistically significant when comparing content, lexis and mechanics with discourse, form[4] and other. On the other hand, the children in the LTG seemed to focus on content, form, lexis and mechanics to a statistically higher extent than on discourse.

[4] Although there are not statistically significant differences between mechanics and form, the effect size is large enough to be highlighted

Table 19. Results for within-group differences across categories at Stage 2, Cycle 2

Group	Categories	*p*	*d*	*95% Confidence Interval*	
				Lower bound	**Upper bound**
CG	Content-Mechanics	.018*	-0.57	-3.87	1.87
	Content-Lexis	1.00	0.47	-1.80	3.13
	Content-Form	.005*	-0.19	-2.86	2.20
	Content-Discourse	1.00	0.38	-1.49	2.60
	Content-Other	1.00	0.68	-1.17	2.94
	Mechanics-Lexis	.003*	1.22	0.43	2.90
	Mechanics-Form	1.00	0.38	-1.23	2.56
	Mechanics-Discourse	.048*	1.11	-0.25	3.36
	Mechanics-Other	.006*	1.49	-0.05	3.83
	Lexis-Form	.028*	-0.70	-2.46	0.46
	Lexis-Discourse	1.00	-0.12	-1.54	1.32
	Lexis-Other	1.00	0.31	-1.56	2.01
	Form-Discourse	.468	0.61	-0.37	2.15
	Form-Other	009*	0.92	-0.60	3.04
	Discourse-Other	1.00	0.42	-0.96	1.62
TG	Content-Mechanics	1.00	0.21	-2.23	2.96
	Content-Lexis	1.00	0.10	-2.05	2.41
	Content-Form	.045*	0.82	-0.93	3.65
	Content-Discourse	.017*	1.33	0.25	3.94
	Content-Other	.004*	1.08	-0.13	3.59
	Mechanics-Lexis	1.00	-0.12	-1.30	0.93
	Mechanics-Form	1.00	0.79	-0.71	2.71
	Mechanics-Discourse	.031*	1.49	0.09	3.36
	Mechanics-Other	.047*	1.13	-0.39	3.12
	Lexis-Form	.014*	0.87	-0.14	2.50
	Lexis-Discourse	.001*	1.52	0.62	3.20
	Lexis-Other	.031*	1.19	-0.07	3.16
	Form-Discourse	.744	0.42	-0.41	1.87
	Form-Other	1.00	0.35	-1.28	2.01
	Discourse-Other	1.00	-0.41	-1.53	0.80
LTG	Content-Mechanics	1.00	0.31	-2.22	3.22
	Content-Lexis	1.00	0.31	-1.84	2.84
	Content-Form	1.00	0.13	-2.20	2.60
	Content-Discourse	.003*	1.01	-0.44	3.44
	Content-Other	.025*	0.62	-0.95	2.95
	Mechanics-Lexis	1.00	0	-1.17	1.17
	Mechanics-Form	.524	-0.28	-2.10	1.50
	Mechanics-Discourse	.026*	1.05	-0.71	2.71
	Mechanics-Other	.1.00	0.44	-1.34	2.34
	Lexis-Form	1.00	-0.30	-1.68	1.08
	Lexis-Discourse	.013*	1.13	-0.35	2.35
	Lexis-Other	1.00	0.46	-1.19	2.19
	Form-Discourse	.024*	1.56	0.11	2.49
	Form-Other	1.00	0.76	-0.93	2.53
	Discourse-Other	1.00	-0.54	-1.72	0.72

Among the features that the children in the CG verbalized the most we find spelling and punctuation issues. The learners also struggled with verb tenses, the 3rd person singular morpheme, the use of prepositions or the dropping of subjects. They also experienced problems with stylistic issues such as explicitness (CHI3: 'And we can eh… don't say 'person' and say 'one boy') or the use of synonyms in order to avoid repetition (CHI6: '… The dog see the cientific and…' And Michael, that way we don't say 'cientific' again). The TG was predominantly engaged in episodes related to content, spelling, punctuation, verbs, nouns, adverbs, verb tenses, possessive pronouns and prepositions. Similarly, the LTG observed differences in content, spelling, punctuation, verbs, nouns, verb forms and possessive pronouns (see Appendix 30).

When the three treatments were compared, post-hoc tests located differences in content and lexis between the CG and the TG, suggesting once again that the model had a significant effect on the children's attention to lexical and content-related features. Contrary to expectations, however, the LTG did not exhibit any difference when compared to the other groups as to their focus of attention. Table 20 below illustrates these findings.

Table 20. Results for between-group differences at Stage 2, Cycle 2

Categories per group	**POST-HOC**				
	Group	***p***	***d***	***95% Confidence Interval***	
				Lower bound	**Upper bound**
CG-TG-LTG: Content	CG-TG	.008*	-0.29	-3.66	0.83
	TG-LTG	1.00	0.18	-1.35	3.02
	CG-LTG	1.00	-0.12	-2.88	1.72
CG-TG-LTG: Lexis	CG-TG	.006*	-0.63	-3.31	-0.48
	TG-LTG	.124	0.39	-0.22	2.53
	CG-LTG	.600	-0.24	-2.19	0.70

In summary, at Stage 1, no between-group differences were observed in the total number of CREs and LREs produced in Cycle 1, nor between groups as far as types of features are concerned. The three groups focused their attention mostly on lexis, form and mechanics. At Stage 2 of Cycle 1, we can observe a first difference between the CG and the treatment groups in that the latter (i) noticed significantly more features than the CG when comparing their drafts with the model and (ii) attended to lexis and content to

a significantly higher extent than other categories. By contrast, the CG focused mainly on form and mechanics. As a result, these findings suggest a role for models in promoting engagement in a higher number of episodes as well as noticing of lexical and content-related episodes, while self-correction seems to foster children's attention to grammar and mechanical aspects.

At the beginning of Cycle 2, the LTG produced significantly fewer episodes than their counterparts. The children in the CG and the TG continued experiencing mostly mechanical, lexical and formal problems. Nevertheless, apart from these three categories, the LTG also allocated significantly more attentional resources to discourse as opposed to Cycle 1. As for inter-subject differences, the TG attended more to CREs than the LTG, and the CG focused significantly more on mechanics than their classmates in the LTG. At Stage 2, the CG and the LTG engaged in significantly fewer episodes than the TG, although the LTG still produced significantly more episodes than the CG. Additionally, the three groups showed differences in their focus of attention: On the one hand, while correcting their texts, the CG's focus was entirely on mechanics but also on form over lexis and content. For their part, the children in the TG attended mostly to content and lexis, as they did in Cycle 1, but also to mechanics. Surprisingly, the dyads in the LTG identified formal, mechanical, lexical and content-related features, a wider range of categories than in Cycle 1. Nevertheless, the only statistically significant difference found across groups is that between the TG and the CG in terms of content and lexis.

Finally, research question 1c asked whether there exist within-group differences across cycles regarding frequency, type and outcome of the episodes produced in oral interaction. In relation to frequency, we found some differences, as Table 21 shows. The CG significantly decreased the proportion of episodes noticed at the composing stage of Cycle 2 with respect to those noticed in Cycle 1. Likewise, the LTG produced a significantly smaller proportion of PFNs, while no differences were observed for the TG. At Stage 2, no differences were found in the total amount of features identified by the CG and the TG separately, but the LTG's production of episodes did decrease significantly in Cycle 2, as in Stage 1.

Table 21. Results for across-cycle differences in the proportion of PFNs and FNs produced by each group

Stage	Groups					
	CG		TG		LTG	
	z	*p*	*z*	*p*	*z*	*p*
Stage 1	3.49	< .000*	0.83	.401	4.89	< .000*
Stage 2	0.12	.900	1.43	.150	5.70	< .000*

Regarding changes in the types of features, as can be seen from the data in Table 22, the results obtained from analyzing differences in the categories noticed across cycles are somewhat counterintuitive. After the long-term treatment and the significant differences observed between groups in Cycle 2, we would expect a notable gain in the noticing of content-related and lexical episodes on the part of the LTG. Nonetheless, not only is there no such an increase, but we observe a significant decrease in these mean values at both stages 1 and 2 of Cycle 2. Interestingly, mechanical and grammatical LREs also underwent a statistically significant drop at Stage 1 despite the LTG having attended to these two categories to a significant extent. Like the LTG, the lexical features noticed by the TG decreased at Stage 1, but the proportion of mechanical LREs increased at Stage 2. As for the dyads in the self-correction group, no differences were detected in any of the stages.

Table 22. Results for differences in the types of features noticed across cycles

Groups	Categories		*p*	*d*	*95% Confidence Interval*	
					Lower bound	**Upper bound**
CG	Stage 1	Content	.233	0.41	-0.45	1.79
	Stage 2		.254	-0.21	-2.76	0.76
	Stage 1	Mechanics	.288	-0.36	-3.86	1.19
	Stage 2		.136	-0.28	-2.08	0.30
	Stage 1	Lexis	.063	0.65	-0.10	3.66
	Stage 2		.898	0.02	-1.65	1.87
	Stage 1	Form	.581	0.19	-1.78	3.12
	Stage 2		.150	0.27	-0.34	2.12
	Stage 1	Discourse	.333	0.33	-0.36	1.03
	Stage 2		.389	0.16	-0.45	1.11
	Stage 1	Other	.352	-0.32	-0.70	0.26
	Stage 2		.308	-0.19	-0.99	0.32
TG	Stage 1	Content	.053	0.61	-0.01	2.01
	Stage 2		.139	0.28	-0.41	2.77
	Stage 1	Mechanics	.871	0.05	-2.10	2.47
	Stage 2		.001*	-0.67	-2.98	-0.83
	Stage 1	Lexis	.008*	0.86	0.66	4.06
	Stage 2		.251	0.21	-0.68	2.50
	Stage 1	Form	.284	0.33	-1.03	3.40
	Stage 2		.740	-0.06	-1.30	0.93
	Stage 1	Discourse	.557	-0.18	-0.81	0.45
	Stage 2		.125	0.29	-0.16	1.25
	Stage 1	Other	.672	-0.13	-0.53	0.34
	Stage 2		.756	0.06	-0.50	0.69
LTG	Stage 1	Content	.012*	0.85	0.34	2.46
	Stage 2		.027*	0.43	0.23	3.57
	Stage 1	Mechanics	.034*	0.70	0.20	5.00
	Stage 2		.157	-0.27	-1.93	0.33
	Stage 1	Lexis	.048*	0.66	0.02	3.58
	Stage 2		.005*	0.56	0.83	4.17
	Stage 1	Form	.011*	0.87	0.78	5.42
	Stage 2		.862	-0.03	-1.27	1.07
	Stage 1	Discourse	.358	-0.30	-0.96	0.36
	Stage 2		.178	0.25	-0.24	1.24
	Stage 1	Other	.657	-0.14	-0.56	0.36
	Stage 2		.333	-0.18	-0.92	0.32

On the subject of across-cycle differences in the outcome of episodes, the three groups were subjected to comparison at Stage 1. However, only the CREs and LREs produced by the CG were considered for the Stage 2 comparison, as the pairs in the treatment

groups did not have to solve any of the features noticed in the models. Table 23 presents the frequency and proportions of the resolution of episodes according to whether they were resolved (correct or incorrect) or unresolved (addressed or ignored) (see Appendix 35 for a breakdown of the outcome of episodes per subcategories).

Table 23. Frequency and proportions of the resolution of episodes noticed at Stages 1 and 2 of Cycle 1 and 2

Group	**Resolution**				
	Correct	**Incorrect**	**Addressed**	**Ignored**	**Total**
Cycle 1					
Stage 1					
CG	100	39	7	8	154
%	65%	25%	5%	5%	100%
TG	136	36	8	6	186
%	73%	19%	4%	3%	100%
LTG	125	34	7	7	173
%	72%	20%	4%	4%	100%
Stage 2					
CG	33	16	2	0	51
%	65%	31%	4%	0%	100%
Cycle 2					
Stage 1					
CG	100	26	4	3	133
%	75%	20%	3%	2%	100%
TG	86	46	4	1	137
%	63%	33%	3%	1%	100%
LTG	72	12	3	1	88
%	82%	14%	3%	1%	100%
Stage 2					
CG	38	15	4	2	59
%	65%	25%	7%	3%	100%

As detailed in Table 24, we obtained statistically significant intra-subject differences in the outcome of the episodes generated. The CG and the LTG significantly increased the number of correctly resolved episodes in Cycle 2, while the TG underwent a significant decrease. Not only did the TG resolve fewer episodes satisfactorily, but they also produced a higher number of incorrect ones, while the LTG managed to reduce the number of incorrectly resolved LREs and CREs. The three groups seemed to reduce the number of ignored episodes, and no differences were found for the addressed episodes nor for Stage 2.

Table 24. Results for across-cycle differences in the resolution of PFNs and FNs

	Resolution							
	Correct		**Incorrect**		**Addressed**		**Ignored**	
	z	*p*	*z*	*p*	*z*	*p*	*z*	*p*
Stage 1								
CG	2.94	.003*	1.80	.070	1.11	.263	2.45	.014*
TG	2.91	.003*	4.10	< .000*	1.11	.263	3.99	< .000*
LTG	3.26	.001*	2.31	.021*	0.46	.644	3.61	< .000*
Stage 2								
CG	0.04	.964	0.97	.329	0.81	.416	1.33	.180

6.1.1.1 Summary of the main findings for research question 1

Research question 1a: What features of language do they notice when composing a narrative text?

Stage 1-Cycle 1

- No differences between groups were found in the total number of features noticed while writing a text in collaboration.
- Each of the three groups attended mostly to lexis, form and mechanics.
- No between-group differences were found for any of the categories analyzed. This suggests that, at least initially, the participants assigned to either of the three groups were similar with respect to the problems noticed.

Stage 1-Cycle 2

- The children in the LTG verbalized significantly fewer episodes than their counterparts.
- As for differences across categories, the CG and the TG continued experiencing mostly mechanical, lexical and formal problems. Apart from these three categories, the LTG also seemed to allocate significantly more attentional resources to discourse as opposed to Cycle 1.
- After being subjected to different treatments, the participants in the three feedback conditions continued focusing on lexis, form, discourse and other issues to a similar extent during the composing stage. Nevertheless, the

problems that the TG encountered with content were found to be greater than those of the LTG. In like fashion, the self-correction group was found to show more interest in mechanics than the children in the LTG.

Research question 1b: What do they notice when comparing their written texts with model texts or when correcting themselves?

Stage 2-Cycle 1

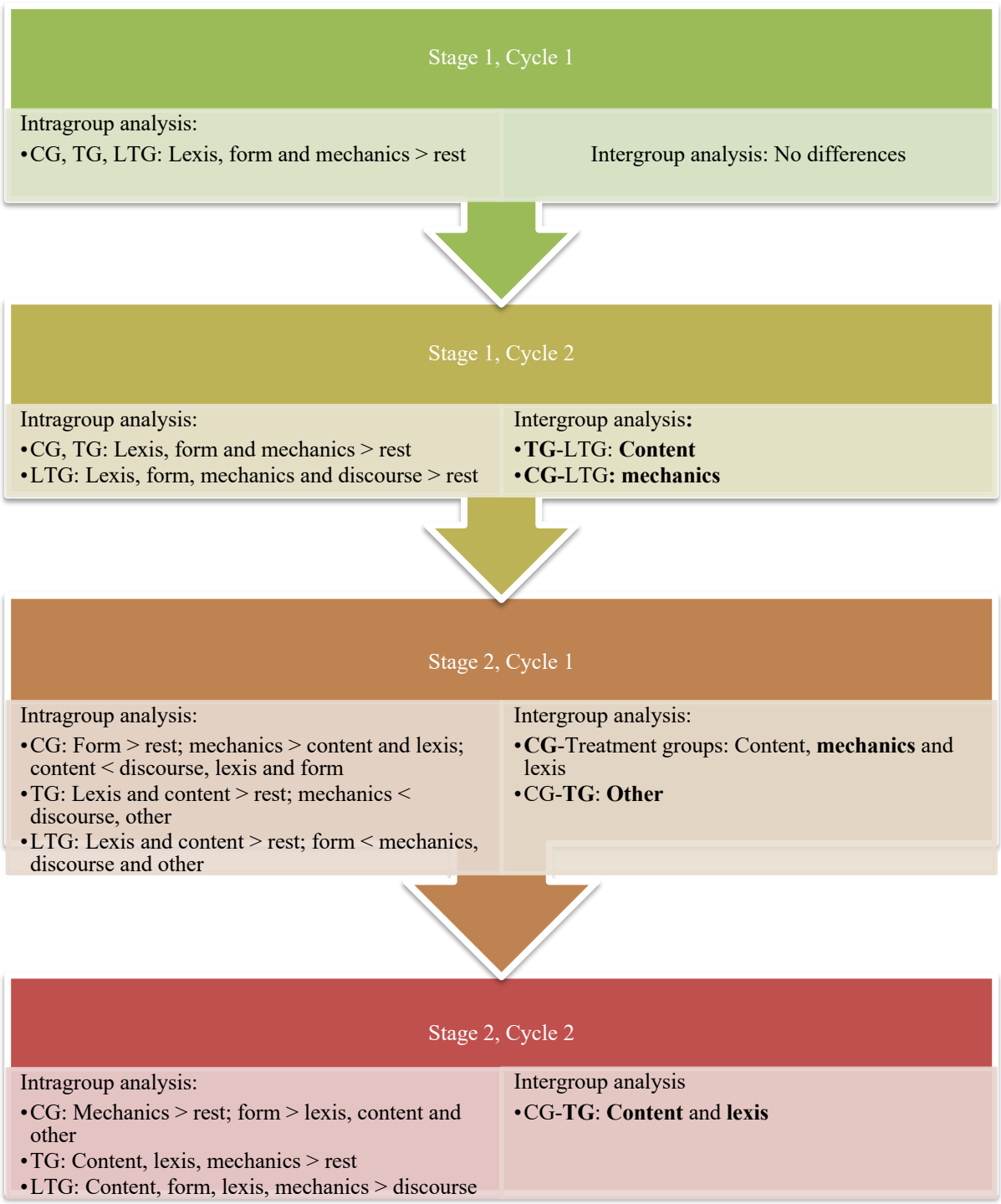
- The treatment groups noticed a significantly higher number of features than the CG.
- The treatment groups noticed significantly more lexical and content-related features than any other category and the difference was also significant with respect to the CG's noticing of the same features.
- The CG focused mainly on form and mechanics, the difference of the latter also being significant with respect to the treatment groups' noticing of mechanics. Therefore, the model text seemed to prompt a higher number of lexical and content-related episodes, while self-correction triggered a predominant focus on form and mechanics.

Stage 2-Cycle 2

- The TG identified a statistically higher number of episodes than the CG and the LTG, and the difference between the CG and the LTG was also significant.
- The CG kept attending mostly to form and mechanics, the TG focused on content, lexis and mechanics, and the learner pairs in the LTG were substantially engaged in formal, lexical, content-related and mechanical LREs.
- As for differences across feedback conditions, differences were observed in the scores for content and lexis between the CG and the TG. Contrary to expectations, however, the LTG did not exhibit any difference when compared to the other groups as to their focus of attention.

Figure 3 provides a summary of the results for research questions 1a and 1b.

Figure 3. Summary of research questions 1a and 1b



Research question 1c: Are there any across-cycle differences within groups regarding frequency, type and outcome of the episodes produced in oral interaction?

- In relation to frequency, the comparison of the first stages revealed that both the CG and the LTG produced a significantly lower proportion of episodes in Cycle 2, as opposed to Cycle 1. When the comparison stages were contrasted, we found that the LTG engaged in significantly fewer episodes as compared with Cycle 1.
- Regarding types of features, the number of grammatical and mechanical episodes at Stage 1 and of content-related and lexical aspects at stages 1 and 2 of Cycle 2 decreased in the LTG. The number of lexical features noticed at Stage 1 decreased in the TG but the proportion of mechanical LREs at Stage 2 increased. No differences were detected in the case of the self-correction group.
- As for the resolution of episodes, the number of correctly resolved episodes on the part of the TG decreased from Cycle 1 to Cycle 2, while the incorrect ones increased. For their part, the CG and the LTG produced a higher number of correctly resolved episodes in Cycle 2, and the LTG reduced the incorrect ones. The three groups seemed to reduce the number of ignored episodes, and no differences were found for the addressed episodes nor for Stage 2.

6.1.2 Results for research question 2

The second research question asked the following:

2. How is the children's reported noticing at the composition and comparison stages related to their revised texts? (Impact of oral production on written production).

This question is concerned with the number and types of revisions made by the learner pairs to their drafts and the relationship of these incorporations to both the writing (Stage 1) and comparison (Stage 2) stages. To this end, the incorporations identified in the dyads' texts were classified as (i) those that originated only at Stage 1, (ii) those that surfaced at Stage 2 from either comparison with the models or self-correction, (iii) those noticed by the children at both Stages 1 and 2, and (d) those features which had not been noticed or verbalized by the children but that could be traced to the treatment they had received (FUIs). Table 25 reports summary statistics of the features incorporated into the revised texts produced at Stage 3 of Cycle 1.

Table 25. Descriptive statistics of features incorporated at Stage 3 of Cycle 1

	CG (*N*=9)			**TG (*N*=11)**			**LTG (*N*=10)**		
	FO	*Mean*	*SD*	*FO*	*Mean*	*SD*	*FO*	*Mean*	*SD*
CREs	1	0.11	0.33	18	1.64	0.92	19	1.9	1.37
Mechanics	4	0.44	0.73	7	0.64	0.67	9	0.9	0.88
Lexis	4	0.44	0.73	26	2.36	1.12	27	2.7	0.95
Form	10	1.11	1.76	12	1.09	1.04	12	1.2	0.79
Discourse	0	0	0	3	0.27	0.47	9	0.9	0.99
Other	0	0	0	1	0.09	0.30	0	0	0

Table 26 presents the frequencies and proportions of the features incorporated (FNIs and FUIs) into the children's writing in relation to the stage or stages in which they emerged. Overall, the features incorporated by the children in the CG were mostly formal (53%), while the treatment groups made mostly lexical (39% for TG, 36% for

LTG) and, to a lesser extent, content-related (27% for TG, 25% for LTG) revisions to their drafts at Stage 3 of Cycle 1. Most of these lexical and content-related incorporations from the model either had not been previously reported (46% and 67% for TG, 44% and 32% for LTG) or had been identified at Stage 2 (46% and 17%, 33% and 37%, respectively). As a matter of fact, most of the overall revisions made by the children in the treatment groups originated at the comparison stage (36% for TG, 39% for LTG) as opposed to the CG whose incorporations were mostly identified at Stage 1 (58%). Another interesting aspect we can observe from this table is that a high proportion of the revisions made to the treatment groups' first texts (49% and 34%) had not been verbalized, as compared to the CG whose incorporations had all been noticed, since they had not received any input they could learn from.

Table 26. Frequency and proportion of features incorporated at Stage 3 of Cycle 1 across stages and feedback types

	CG					TG					LTG				
Episodes	S1[1]	S2[2]	Stages 1+2	FUI[3]	**Total**	S1	S2	Stages 1+2	FUI	**Total**	S1	S2	Stages 1+2	FUI	**Total**
Content	0 (0%)	1 (100%)	0 (0%)	0 (0%)	1 (5%)	2 (11%)	3 (17%)	1 (6%)	12 (67%)	18 (27%)	3 (16%)	7 (37%)	3 (16%)	6 (32%)	19 (25%)
Mechanics	2 (50%)	1 (25%)	1 (25%)	0 (0%)	4 (21%)	1 (14%)	1 (14%)	0 (0%)	5 (71%)	7 (10%)	2 (22%)	3 (33%)	0 (0%)	4 (44%)	9 (12%)
Lexis	3 (75%)	0 (0%)	1 (25%)	0 (0%)	4 (21%)	1 (4%)	12 (46%)	1 (4%)	12 (46%)	26 (39%)	4 (15%)	9 (33%)	2 (7%)	12 (44%)	27 (36%)
Form	6 (60%)	2 (20%)	2 (20%)	0 (0%)	10 (53%)	2 (17%)	6 (50%)	0 (0%)	4 (33%)	12 (18%)	2 (17%)	6 (50%)	1 (8%)	3 (25%)	12 (16%)
Discourse	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	1 (33%)	1 (33%)	1 (33%)	0 (0%)	3 (4%)	3 (33%)	5 (56%)	0 (0%)	1 (11%)	9 (12%)
Other	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	1 (100%)	0 (0%)	0 (0%)	1 (1%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)
Total	11 58%	4 21%	4 21%	0 0%	**19**	7 10%	24 36%	3 4%	33 49%	**67**	14 18%	30 39%	6 8%	26 34%	**76**

[1] Stage 1; [2] Stage 2; [3] Features unnoticed and incorporated

Regarding proportion of incorporations, the binomial test did not show any statistically significant differences between the treatment groups (z = 1.01, p = .313), but did find differences between them and the CG (z = 6.04, p = < .000 for TG; z = 6.96, p = < .000 for LTG).

Statistical tests were also run to look for within-group and between-group differences with respect to categories. Sphericity was verified for Categories (X^2 = 34.90; p = .002) and for the interaction between Categories and Cycles (X^2 = 34.29; p = .002). The assumption of sphericity was violated, so the Greenhouse-Geisser correction was used. Table 27 below shows the results of the mixed ANOVA for Stage 3. As illustrated, statistical tests revealed that there is a main effect for Categories ($F_{(3.58,95.77)}$ = 41.87; p = < .001; η_p^2 = 0.25), Group ($F_{(2,27)}$ = 52.16; p = < .001; η_p^2 = 0.16) and for Categories x Group ($F_{(7.09,95.77)}$ = 8.84; p = < .001; η_p^2 = 0.1) and Categories x Cycles x Group interactions ($F_{(7.35,99.23)}$ = 7.89; p = .049; η_p^2 = 0.13).

Table 27. Mixed-model ANOVA results for Stage 3

	Stage 3				
	***df*1**	***df*2**	***F***	***p***	η_p^2
Categories	3.58	95.77	41.87	< .001*	0.25
Categories * Group	7.09	95.77	8.84	< .001*	0.1
Cycles	1	27	0.83	.370	0.00
Cycles * Group	2	27	1.42	.259	0.01
Categories * Cycles	3.68	99.23	2.4	.060	0.02
Categories * Cycles * Group	7.35	99.23	7.89	.049*	0.13
Inter-subjects					
Group	2	27	52.16	< .001*	0.16

Further post-hoc tests revealed significant differences for both the TG and the LTG. As observed in the table below, both treatment groups largely coincide in the types of categories most significantly incorporated after being exposed to models, namely, content and lexis, the only difference being that lexis is significantly different from the rest of the categories, while content is not different from form. These analyses lend support to the view that models had a statistically significant impact on the incorporation of lexical, but also of content-related features and, while formal aspects did not show any statistically homogeneous power, we cannot overlook the fact that the children also incorporated a high number of grammatical items. For their part, the pairs

in the self-correction group made revisions related mostly to grammar issues (as can be seen in Table 26), although no statistically significant differences were found.

Table 28. Results for across-category differences in the incorporation of features at Stage 3 of Cycle 1 per group

Group	Categories	*p*	*d*	*95% Confidence Interval*	
				Lower bound	**Upper bound**
TG	Content-Mechanics	.002*	0.44	-0.35	2.35
	Content-Lexis	1.00	-0.32	-2.07	0.62
	Content-Form	1.00	0.20	-1.03	2.12
	Content-Discourse	.004*	0.77	0.32	2.41
	Content-Other	<.000*	0.98	0.62	2.47
	Mechanics-Lexis	.001*	-0.84	-2.93	-0.52
	Mechanics-Form	1.00	-0.17	-2.05	1.14
	Mechanics-Discourse	1.00	0.22	-0.60	1.33
	Mechanics-Other	.023*	0.41	-0.24	1.33
	Lexis-Form	.043*	0.60	0.02	2.52
	Lexis-Discourse	<.000*	1.23	1.09	3.09
	Lexis-Other	<.000*	1.44	1.35	3.20
	Form-Discourse	.013*	0.37	-0.48	2.12
	Form-Other	.046*	0.49	-0.19	2.19
	Discourse-Other	1.00	0.18	-0.42	0.79
LTG	Content-Mechanics	.017*	0.42	-0.41	2.41
	Content-Lexis	1.00	-0.33	-2.21	0.61
	Content-Form	1.00	0.25	-0.95	2.35
	Content-Discourse	.031*	0.54	-0.10	2.10
	Content-Other	<.000*	1.15	0.93	2.87
	Mechanics-Lexis	.001*	-0.84	-3.07	-0.53
	Mechanics-Form	1.00	-0.11	-1.97	1.37
	Mechanics-Discourse	1.00	0.00	-1.01	1.01
	Mechanics-Other	.024*	0.64	0.07	1.73
	Lexis-Form	.015*	0.67	0.19	2.81
	Lexis-Discourse	<.000*	1.01	0.75	2.85
	Lexis-Other	<.000*	1.64	1.73	3.67
	Form-Discourse	1.00	0.13	-1.07	1.67
	Form-Other	.049*	0.57	-0.05	2.45
	Discourse-Other	.002*	0.83	0.26	1.54

Comparisons across groups support the results stated above. Post-hoc analyses found a statistically significant effect for content and lexis between the treatment groups and the CG. This indicates that the use of models as a feedback technique leads to a statistically significant improvement when it comes to incorporating vocabulary and ideas in

comparison to having the children correct their own texts. Table 29 presents these results.

Table 29. Results for between-group differences in the incorporation of features at Stage 3 of Cycle 1 per category

Categories per group	**Group**	***p***	***d***	**POST-HOC**	
				95% Confidence Interval	
				Lower bound	**Upper bound**
CG-TG-LTG: Content	CG-TG	.006*	-0.63	-2.66	-0.39
	TG-LTG	1.00	-0.11	-1.37	0.84
	CG-LTG	.002*	-0.72	-2.95	-0.63
CG-TG-LTG: Lexis	CG-TG	<.000*	-0.81	-3.02	-0.82
	TG-LTG	1.00	-0.15	-1.41	0.73
	CG-LTG	<.000*	-0.93	-3.38	-1.13

Moving now to Cycle 2, Table 30 details the descriptive statistics of the features incorporated at Stage 3 of Cycle 2.

Table 30. Descriptive statistics of features incorporated at Stage 3 of Cycle 2

	CG (*N*=9)			**TG (*N*=11)**			**LTG (*N*=10)**		
	FO	*Mean*	*SD*	*FO*	*Mean*	*SD*	*FO*	*Mean*	*SD*
CREs	2	0.22	0.44	22	2	1	26	2.6	1.08
Mechanics	5	0.56	0.73	17	1.55	1.29	16	1.6	0.97
Lexis	2	0.22	0.44	21	1.91	1.14	33	3.3	1.34
Form	2	0.22	0.44	6	0.55	0.69	14	1.4	1.08
Discourse	1	0.11	0.33	7	0.64	0.68	7	0.7	0.82
Other	0	0	0	0	0	0	1	0.1	0.32

Table 31 illustrates the breakdown of the revisions made at Stage 3 after five months receiving models, self-correcting or following no treatment at all. This table is quite revealing in several ways. To start with, we can now observe differences between the three groups. As has been the case until now, significant differences were found

between the model groups and the CG regarding proportion of incorporations (z = 7.55, p = < .000 for TG; z = 9.72, p = 0 for LTG), and for the first time, the binomial test also revealed differences between the TG and the LTG (z = 2.52, p = .011). Therefore, a positive correlation was found between a long-term treatment with models and number of incorporations. Additionally, the pairs in the CG incorporated approximately the same number of features for each category, with a slight blip in mechanics, and actually, here again, no statistically significant differences were observed across categories. As usual, most of the revisions had originated at the moment of writing their first draft (58%) and no unnoticed and incorporated features were found. On the other hand, we can observe that the items incorporated by the TG were mainly related to CREs (30%), vocabulary (29%), but also mechanics (23%), which coincide with the episodes this group most attended to at Stage 2 (see Table 18). Once again, most incorporations either had been identified at Stage 2 (40%) or had not been previously verbalized (34%). In the case of the LTG, changes to lexis and content accounted for 61% of the total revisions made, originating mainly at both the initial stage (21%) and the comparison stage (18%), and FUIs accounted for more than half (58%) of the total incorporations.

Table 31. Frequency and proportion of features incorporated at Stage 3 of Cycle 2 across stages and feedback types

	CG					TG					LTG				
Episodes	S1	S2	Stages 1+2	FUI	**Total**	S1	S2	Stages 1+2	FUI	**Total**	S1	S2	Stages 1+2	FUI	**Total**
Content	1 (50%)	0 (0%)	1 (50%)	0 (0%)	2 (17%)	3 (14%)	5 (23%)	3 (14%)	11 (50%)	22 (30%)	4 (15%)	3 (11%)	2 (7%)	17 (65%)	26 (27%)
Mechanics	3 (60%)	1 (20%)	1 (20%)	0 (0%)	5 (42%)	3 (18%)	8 (47%)	2 (12%)	4 (24%)	17 (23%)	6 (37%)	4 (25%)	0 (0%)	6 (37%)	16 (16%)
Lexis	1 (50%)	1 (50%)	0 (0%)	0 (0%)	2 (17%)	2 (10%)	11 (52%)	1 (5%)	7 (33%)	21 (29%)	6 (18%)	7 (21%)	0 (0%)	20 (60%)	33 (34%)
Form	2 (100%)	0 (0%)	0 (0%)	0 (0%)	2 (17%)	2 (33%)	3 (50%)	1 (17%)	0 (0%)	6 (8%)	3 (21%)	3 (21%)	1 (7%)	7 (50%)	14 (14%)
Discourse	0 (0%)	1 (100%)	0 (0%)	0 (0%)	1 (8%)	2 (29%)	2 (29%)	0 (0%)	3 (43%)	7 (10%)	1 (14%)	1 (14%)	0 (0%)	5 (71%)	7 (7%)
Other	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	1 (100%)	1 (1%)
Total	7 58%	3 25%	2 17%	0 0%	**12**	12 16%	29 40%	7 10%	25 34%	**73**	20 21%	18 18%	3 3%	56 58%	**97**

Results for across-category differences in the incorporation of features at Stage 3 of Cycle 2 are shown in Table 32. Post-hoc analyses showed differences across categories for the model groups. More specifically, lexis, content and mechanics were found to be significantly different from form, discourse and other in the case of the TG. Concerning the LTG, statistical tests supported the group's dominant focus on lexis and content, finding significant differences between these two categories and the remainder. Furthermore, post-hoc tests also indicated an effect for form and mechanics over discourse and other issues. These results are also in line with those obtained at Stage 2 of Cycle 2, in which the LTG appeared to attend to a wider range of categories than ever before.

Table 32. Results for across-category differences in the incorporation of features at Stage 3 of Cycle 2 per group

Group	Categories	*p*	*d*	*95% Confidence Interval*	
				Lower bound	**Upper bound**
TG	Content-Mechanics	1.00	0.19	-0.98	1.89
	Content-Lexis	1.00	0.05	-1.05	1.23
	Content-Form	.004*	0.76	0.33	2.58
	Content-Discourse	.005*	0.75	0.30	2.43
	Content-Other	<.000*	1.34	1.12	2.88
	Mechanics-Lexis	1.00	-0.16	-1.67	0.94
	Mechanics-Form	.049*	0.58	-0.02	2.02
	Mechanics-Discourse	.034*	0.44	-0.30	2.12
	Mechanics-Other	<.000*	0.93	0.57	2.52
	Lexis-Form	.004*	0.77	0.32	2.41
	Lexis-Discourse	.015*	0.67	0.16	2.39
	Lexis-Other	<.000*	1.10	0.88	2.93
	Form-Discourse	1.00	-0.05	-1.11	0.93
	Form-Other	.420	0.42	-0.21	1.30
	Discourse-Other	.098	0.54	-0.06	1.33
LTG	Content-Mechanics	.028*	0.39	-0.50	2.50
	Content-Lexis	1.00	-0.35	-1.89	0.49
	Content-Form	.044*	0.60	0.02	2.38
	Content-Discourse	<.000*	1.00	0.79	3.01
	Content-Other	<.000*	1.60	1.58	3.42
	Mechanics-Lexis	.007*	-0.73	-3.07	-0.33
	Mechanics-Form	1.00	0.11	-0.87	1.27
	Mechanics-Discourse	.046*	0.42	-0.37	2.17
	Mechanics-Other	.001*	0.86	0.47	2.53
	Lexis-Form	<.000*	1.02	0.80	3.00
	Lexis-Discourse	<.000*	1.31	1.43	3.77
	Lexis-Other	<.000*	1.75	2.13	4.27
	Form-Discourse	.047*	0.38	-0.37	1.77
	Form-Other	<.000*	0.96	0.51	2.09
	Discourse-Other	.199	0.48	-0.13	1.33

Post-hoc analyses also indicated that the groups were different in the types of revisions made, finding a statistically significant effect for content, lexis and form. It can be seen from the data in Table 33 that the treatment groups incorporated significantly more content-related and lexical features than the CG at Stage 3 of Cycle 2. The lexical difference was also found between the model groups, which suggests that the LTG incorporated more vocabulary than their counterparts. As for grammar, the LTG also appeared to make more formal revisions than the CG and TG.

Table 33. Results for between-group differences in the incorporation of features at Stage 3 of Cycle 2 per category

Categories per group	**Group**	***p***	***d***	**POST-HOC**	
				95% Confidence Interval	
				Lower bound	**Upper bound**
CG-TG-LTG: Content	CG-TG	<.000*	-0.80	-2.81	-0.74
	TG-LTG	.418	-0.28	-1.61	0.41
	CG-LTG	<.000*	-1.05	-3.44	-1.32
CG-TG-LTG: Lexis	CG-TG	.005*	-0.64	-2.91	-0.47
	TG-LTG	.018*	-0.55	-2.58	-0.20
	CG-LTG	<.000*	-1.15	-4.33	-1.83
CG-TG-LTG: Form	CG-TG	1.00	-0.17	-1.23	0.58
	TG-LTG	.049*	-0.45	-1.73	0.02
	CG-LTG	.009*	-0.60	-2.10	-0.26

Let us turn now to the two-week delayed post-test. Table 34 below contains descriptive statistics of the features incorporated at Stage 4 of Cycle 1.

Table 34. Descriptive statistics of features incorporated at Stage 4 of Cycle 1

	CG (*N*=9)			**TG (*N*=11)**			**LTG (*N*=10)**		
	FO	*Mean*	*SD*	*FO*	*Mean*	*SD*	*FO*	*Mean*	*SD*
CREs	0	0	0	1	0.09	0.3	4	0.4	0.7
Mechanics	1	0.11	0.33	2	0.18	0.41	5	0.5	0.71
Lexis	0	0	0	7	0.64	0.81	6	0.6	0.7
Form	2	0.22	0.44	2	0.18	0.6	6	0.6	0.7
Discourse	0	0	0	3	0.27	0.65	4	0.4	0.7
Other	0	0	0	0	0	0	0	0	0

Table 35 provides a breakdown of the features incorporated by each group at Stage 4 of Cycle 1 and their tracking. As the dyads were not exposed to the same picture as in previous stages, a considerable reduction in the total amount of incorporations was expected. Closer inspection of the table effectively shows that the total number of revisions made by all participants at Stage 4 dropped to almost a quarter (n = 43) of those incorporated at Stage 3 (n = 162). Nevertheless, we can still draw some conclusions from the data. First, a statistically significant difference between the

treatment groups and the CG (z = 3.18, p = .001 for TG; z = 5.06, p = < .000 for LTG) remains regarding proportion of incorporations, which suggests that the former can still retain and incorporate a high number of features as opposed to the self-correction group, at least in the short term. Secondly, as happened at Stage 3, many of the treatment groups' incorporations from the model either had not been previously verbalized (53% for TG, 44%; 44% for LTG) or had been identified at Stage 2 (40% and 32%, respectively). Conversely, most of the revisions made by the children in the CG originated at the writing stage (67%). Thirdly, in consonance with the previous stage, the CG only carried out formal (67%) and mechanical (33%) revisions while writing for the third time. As for the treatment groups, lexical revisions accounted for 47% of all features incorporated by learners in the TG, while the LTG, apart from lexis (24%), also incorporated formal (24%) and mechanical (20%) aspects of the language.

Table 35. Frequency and proportion of features incorporated at Stage 4 of Cycle 1 across stages and feedback types

	CG					TG					LTG				
Episodes	S1	S2	Stages 1+2	FUI	**Total**	S1	S2	Stages 1+2	FUI	**Total**	S1	S2	Stages 1+2	FUI	**Total**
Content	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	1 (100%)	1 (7%)	1 (25%)	2 (50%)	0 (0%)	1 (25%)	4 (16%)
Mechanics	1 (100%)	0 (0%)	0 (0%)	0 (0%)	1 (33%)	0 (0%)	0 (0%)	0 (0%)	2 (100%)	2 (13%)	0 (0%)	0 (0%)	0 (0%)	5 (100%)	5 (20%)
Lexis	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	4 (57%)	0 (0%)	3 (43%)	7 (47%)	0 (0%)	2 (33%)	0 (0%)	4 (67%)	6 (24%)
Form	1 (50%)	1 (50%)	0 (0%)	0 (0%)	2 (67%)	0 (0%)	1 (50%)	0 (0%)	1 (50%)	2 (13%)	2 (33%)	3 (50%)	1 (17%)	0 (0%)	6 (24%)
Discourse	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	1 (33%)	1 (33%)	1 (33%)	3 (20%)	1 (25%)	1 (25%)	1 (25%)	1 (25%)	4 (16%)
Other	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)
Total	2 67%	1 33%	0 0%	0 0%	**3**	0 0%	6 40%	1 7%	8 53%	**15**	4 16%	8 32%	2 8%	11 44%	**25**

Drawing on statistics to look for significant differences, we first verified sphericity for Categories (X^2 = 49.92; p = < .001) and for the interaction between Categories and Cycles (X^2 = 26.46; p = .023). We used the Greenhouse-Geisser correction since the sphericity assumption was not met. Table 36 provides the results of the mixed ANOVA for Stage 4. Closer inspection of the table shows that there is a main effect for Categories ($F_{(3.19,86.03)}$ = 13.89; p = < .001; η_p^2 = 0.13), Cycles ($F_{(1,27)}$ = 13.03; p = .001; η_p^2 = 0.03), Group ($F_{(2,27)}$ = 16.68; p = < .001; η_p^2 = 0.08), and for the interaction between Categories and Group ($F_{(6.37,86.03)}$ = 2.6; p = .021; η_p^2 = 0.05), Categories and Cycles ($F_{(3.65,98.46)}$ = 6.11; p = < .001; η_p^2 = 0.05) and Categories, Cycles and Group ($F_{(7.29,98.46)}$ = 5.46; p = < .001; η_p^2 = 0.03).

Table 36. Mixed-model ANOVA results for Stage 4

Stage 4					
	***df*1**	***df*2**	***F***	***p***	η_p^2
Categories	3.19	86.03	13.89	< .001*	0.13
Categories * Group	6.37	86.03	2.6	.021*	0.05
Cycles	1	27	13.03	.001*	0.03
Cycles * Group	2	27	1.45	.252	0.00
Categories * Cycles	3.65	98.46	6.11	< .001*	0.05
Categories * Cycles * Group	7.29	98.46	5.46	.048*	0.03
Inter-subjects					
Group	2	27	16.68	< .001*	0.08

Notwithstanding the fact that we can appreciate a trend similar to that of Stage 3 in the types of features incorporated, no significant differences were observed across categories for any of the groups or across groups. Consequently, a clear benefit of models in the short-term incorporation of items at Stage 4 of Cycle 1 could not be identified in this analysis.

Last but not least, let us observe the children's behavior at Stage 4 of Cycle 2. Table 37 provides descriptive statistics of the features incorporated at this stage.

Table 37. Descriptive statistics of features incorporated at Stage 4 of Cycle 2

	CG (*N*=9)			**TG (*N*=11)**			**LTG (*N*=10)**		
	FO	*Mean*	*SD*	*N*	*Mean*	*SD*	*N*	*Mean*	*SD*
CREs	1	0.11	0.33	6	0.55	0.82	5	0.5	0.71
Mechanics	3	0.33	0.71	8	0.73	0.91	8	0.8	0.92
Lexis	4	0.44	0.73	13	1.18	0.87	25	2.5	1.27
Form	1	0.11	0.33	2	0.18	0.41	6	0.6	0.7
Discourse	0	0	0	4	0.36	0.67	7	0.7	1.06
Other	0	0	0	0	0	0	0	0	0

Table 38 below presents the frequency and proportion of features incorporated at Stage 4 of Cycle 2 across stages and feedback types. Like the previous stage, between-group differences in the proportion of incorporations were observed. Namely, the CG incorporated significantly fewer items than the treatment groups ($z = 4.21$, $p = < .000$ for TG; $z = 6.58$, $p = < .000$ for LTG) and, in turn, the LTG made a higher number of revisions than the TG ($z = 2.65$, $p = .007$). In addition, we can see that a low number of lexical (n = 4) and mechanical (n = 3) features were incorporated by the children in the self-correction group into their final texts, but we cannot talk about actual significant differences across categories. As always, the majority of these incorporations had emerged at Stage 1 and no incorporation was a FUI. The TG seemed to retain and incorporate mostly vocabulary (39%) items and like the rest of the stages, most of the incorporations had been verbalized at Stage 2 and also a high number of the revisions had not been reported (39%). The results for the LTG are more straightforward. Lexical revisions accounted for almost half of the incorporations (49%) and this time the revisions were fairly evenly distributed between Stages 1, 2 and 1 + 2, and more than half of the features incorporated (53%) had not been noticed or, at least, verbalized.

Table 38. Frequency and proportion of features incorporated at Stage 4 of Cycle 2 across stages and feedback types

	CG					TG					LTG				
Episodes	S1	S2	Stages 1+2	FUI	**Total**	S1	S2	Stages 1+2	FUI	**Total**	S1	S2	Stages 1+2	FUI	**Total**
Content	0 (0%)	1 (100%)	0 (0%)	0 (0%)	1 (11%)	2 (33%)	1 (17%)	0 (0%)	3 (50%)	6 (18%)	1 (20%)	1 (20%)	0 (0%)	3 (60%)	5 (10%)
Mechanics	2 (67%)	0 (0%)	1 (33)	0 (0%)	3 (33%)	0 (0%)	4 (50%)	2 (25%)	2 (25%)	8 (24%)	3 (37%)	2 (25%)	0 (0%)	3 (37%)	8 (16%)
Lexis	4 (100%)	0 (0%)	0 (0%)	0 (0%)	4 (44%)	1 (7%)	5 (38%)	0 (0%)	7 (54%)	13 (39%)	3 (12%)	4 (16%)	2 (8%)	16 (64%)	25 (49%)
Form	1 (100%)	0 (0%)	0 (0%)	0 (0%)	1 (11%)	0 (0%)	2 (100%)	0 (0%)	0 (0%)	2 (6%)	1 (17%)	1 (17%)	4 (67%)	0 (0%)	6 (12%)
Discourse	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	2 (50%)	1 (25%)	0 (0%)	1 (25%)	4 (12%)	1 (14%)	0 (0%)	1 (14%)	5 (71%)	7 (14%)
Other	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)
Total	7 78%	1 11%	1 11%	0 0%	**9**	5 15%	13 39%	2 6%	13 39%	**33**	9 18%	8 16%	7 14%	27 53%	**51**

Significant differences across categories were found for the model groups, as can be seen from Table 39. Among the differences found for the TG group, lexis turned out to be different from formal and discursive elements, and mechanics was also statistically different from form. Regarding the LTG, post-hoc tests supported the clear predominance of lexis over any other category (see Table 39).

Table 39. Results for across-category differences in the incorporation of features at Stage 4 of Cycle 2 per group

Group	Categories	*p*	*d*	*95% Confidence Interval*	
				Lower bound	**Upper bound**
TG	Content-Mechanics	1.00	-0.11	-1.16	0.79
	Content-Lexis	1.00	-0.29	-1.91	0.64
	Content-Form	1.00	0.26	-0.46	1.19
	Content-Discourse	1.00	0.10	-0.88	1.24
	Content-Other	.177	0.49	-0.10	1.20
	Mechanics-Lexis	1.00	-0.20	-1.82	0.91
	Mechanics-Form	.017*	0.97	-0.33	1.42
	Mechanics-Discourse	1.00	0.23	-0.58	1.31
	Mechanics-Other	.013*	0.51	-0.10	1.56
	Lexis-Form	.028*	0.63	0.07	1.93
	Lexis-Discourse	.046*	0.58	-0.01	1.64
	Lexis-Other	.007*	0.72	0.22	2.14
	Form-Discourse	1.00	-0.15	-0.92	0.55
	Form-Other	1.00	0.22	-0.31	0.67
	Discourse-Other	1.00	0.30	-0.35	1.08
LTG	Content-Mechanics	1.00	-0.17	-1.32	0.72
	Content-Lexis	.001*	-0.88	-3.34	-0.66
	Content-Form	1.00	-0.07	-0.96	0.76
	Content-Discourse	1.00	-0.11	-1.31	0.91
	Content-Other	.386	0.43	-0.18	1.18
	Mechanics-Lexis	.011*	-0.70	-3.14	-0.26
	Mechanics-Form	1.00	0.13	-0.72	1.12
	Mechanics-Discourse	1.00	0.06	-0.89	1.09
	Mechanics-Other	.096	0.54	-0.07	1.67
	Lexis-Form	<.000*	1.14	0.92	2.88
	Lexis-Discourse	<.000*	1.22	0.94	2.66
	Lexis-Other	<.000*	1.46	1.49	3.51
	Form-Discourse	1.00	-0.08	-0.87	0.67
	Form-Other	.013*	0.68	0.08	1.12
	Discourse-Other	.085	0.55	-0.05	1.45

As for differences across treatments, post-hoc tests revealed lexical differences between the LTG and the other two groups, suggesting that a long treatment with models had a significant effect on the children's incorporation of lexical features as opposed to the CG and even to the TG. Table 40 illustrates these findings.

Table 40. Results for between-group differences in the incorporation of features at Stage 4 of Cycle 2 per category

Categories per group	**POST-HOC**				
	Group	***p***	***d***	***95% Confidence Interval***	
				Lower bound	**Upper bound**
CG-TG-LTG: Lexis	CG-TG	.325	-0.30	-1.87	0.40
	TG-LTG	.015*	-0.56	-2.42	-0.22
	CG-LTG	<.000*	-0.83	-3.21	-0.90

On the question of whether differences exist across cycles in the frequency of revisions made by the three groups altogether, binomial tests did not show any significant differences for the treatment groups, as Table 41 shows. However, we did find that the CG incorporated statistically fewer features at Stage 3 of Cycle 2.

Table 41. Results for across-cycle differences in the frequency of incorporations per group

Stage	**Groups**					
	CG		**TG**		**LTG**	
	z	*p*	*z*	*p*	*z*	*p*
Stage 3	2.63	.008*	0.32	.745	1.62	.103
Stage 4	0.59	.549	0.08	.934	0.43	.663

As for variation across cycles in the types of categories incorporated, post-hoc analyses shown in Table 42 revealed no differences in the self-correction condition neither at Stage 3 nor at Stage 4. We observed, however, differences in mechanics for the TG at Stage 3, which largely coincides not only with the rise in mechanical LREs noticed at Stage 2 of Cycle 2, but also with the increase in the incorporation of such features at Stage 3 of Cycle 2. No differences were found between post-tests. Contrary to

expectations, no significant differences across categories were detected for the LTG at Stage 3, but lexis was found to be greater at Stage 4 of Cycle 2.

Table 42. Results for differences in the incorporation of features across cycles

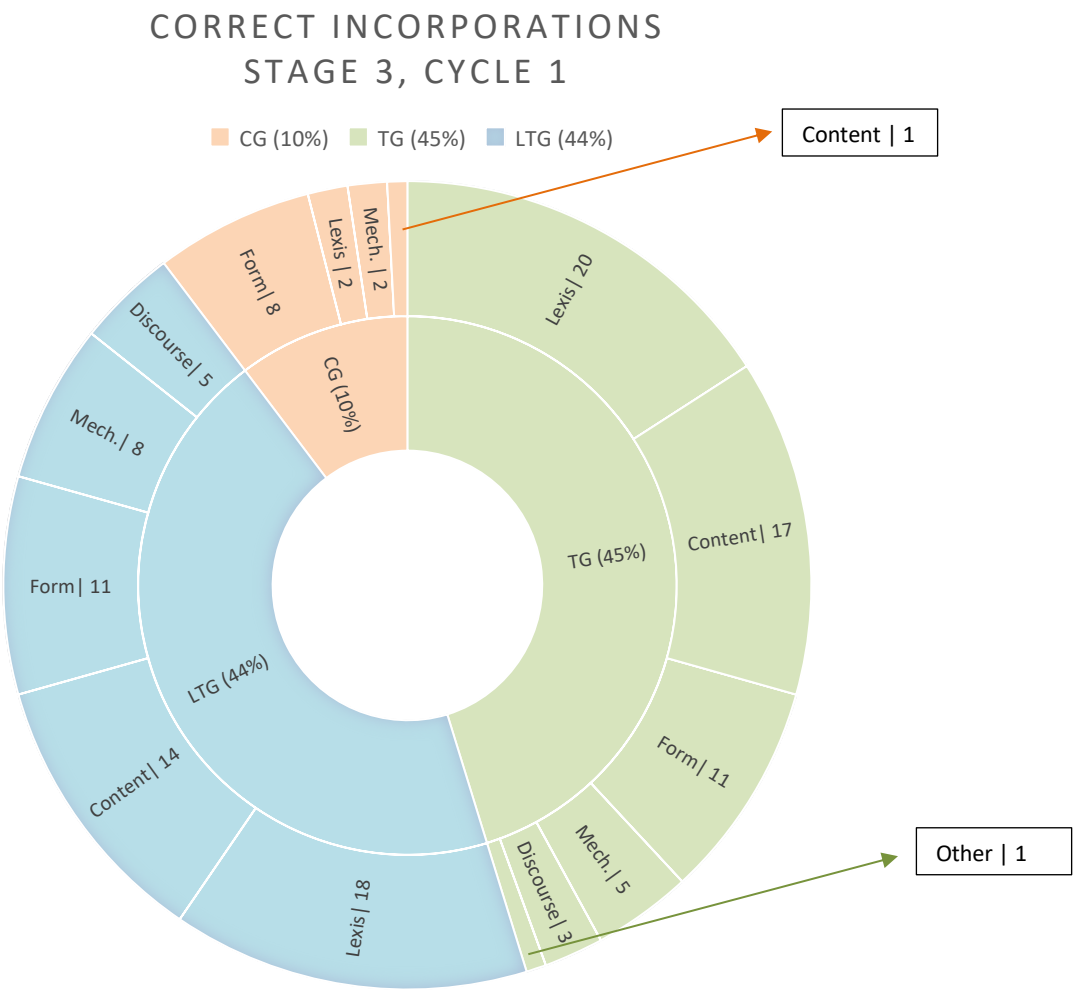
Groups	Categories		*p*	*d*	*95% Confidence Interval*	
					Lower bound	**Upper bound**
TG	Stage 3	Content	.459	-0.14	-1.36	0.63
	Stage 4		.095	-0.32	-0.99	0.08
	Stage 3	Mechanics	.025*	-0.43	-1.70	-0.12
	Stage 4		1.00	-0.31	-1.20	0.11
	Stage 3	Lexis	.331	0.18	-0.49	1.40
	Stage 4		.097	-0.31	-1.20	0.11
	Stage 3	Form	.240	0.22	-0.39	1.48
	Stage 4		1.00	0	-0.54	0.54
	Stage 3	Discourse	.203	-0.24	-0.94	0.21
	Stage 4		.689	-0.07	-0.55	0.37
	Stage 3	Other	.254	0.21	-0.07	0.25
	Stage 4		0	0	0	0
LTG	Stage 3	Content	.179	-0.25	-1.74	0.34
	Stage 4		.720	-0.07	-0.67	0.47
	Stage 3	Mechanics	.094	-0.32	-1.53	0.13
	Stage 4		.379	-0.16	-0.99	0.39
	Stage 3	Lexis	.224	-0.23	-1.59	0.39
	Stage 4		<.000*	-1.04	-2.58	-1.22
	Stage 3	Form	.678	-0.08	-1.18	0.78
	Stage 4		1.00	0	-0.57	0.57
	Stage 3	Discourse	.499	0.13	-0.40	0.80
	Stage 4		.214	-0.23	-0.78	0.18
	Stage 3	Other	.232	-0.22	-0.27	0.07
	Stage 4		0	0	0	0

Besides types and frequency of incorporations, we were also interested in knowing how many of those features were incorporated correctly and whether or not there was an effect of models on accuracy in the short and long run. Graph 1 shows the correct incorporations at Stage 3 of Cycle 1. Out of a total of 162 incorporations, the three groups incorporated a total of 126 (78%) features correctly. As can be seen, 45% (n = 57) and 44% (n = 56) of all correct revisions made by the children correspond to the TG and LTG, respectively, as opposed to the CG whose correct revisions constitute only 10% (n = 13) out of the total number of acceptable incorporations. The difference

between the CG and the treatment groups was found to be significant ($z = 6.18$, $p = <$ .000 for TG; $z = 6.07$, $p = <$ .000 for LTG). If we now calculate the percentage of acceptable answers over the total number of incorporations, we obtain a 68% of correct forms in the case of the CG, 85% for the TG and 74% for the LTG.

Concerning categories, the vast majority of the upgraded revisions made by the treatment groups were content- and lexis-related followed by form, mechanics, discourse and other issues, whereas most of the accurate answers provided by the CG were grammatical (see Appendix 31 for a breakdown of correct and incorrect incorporations per subcategories).

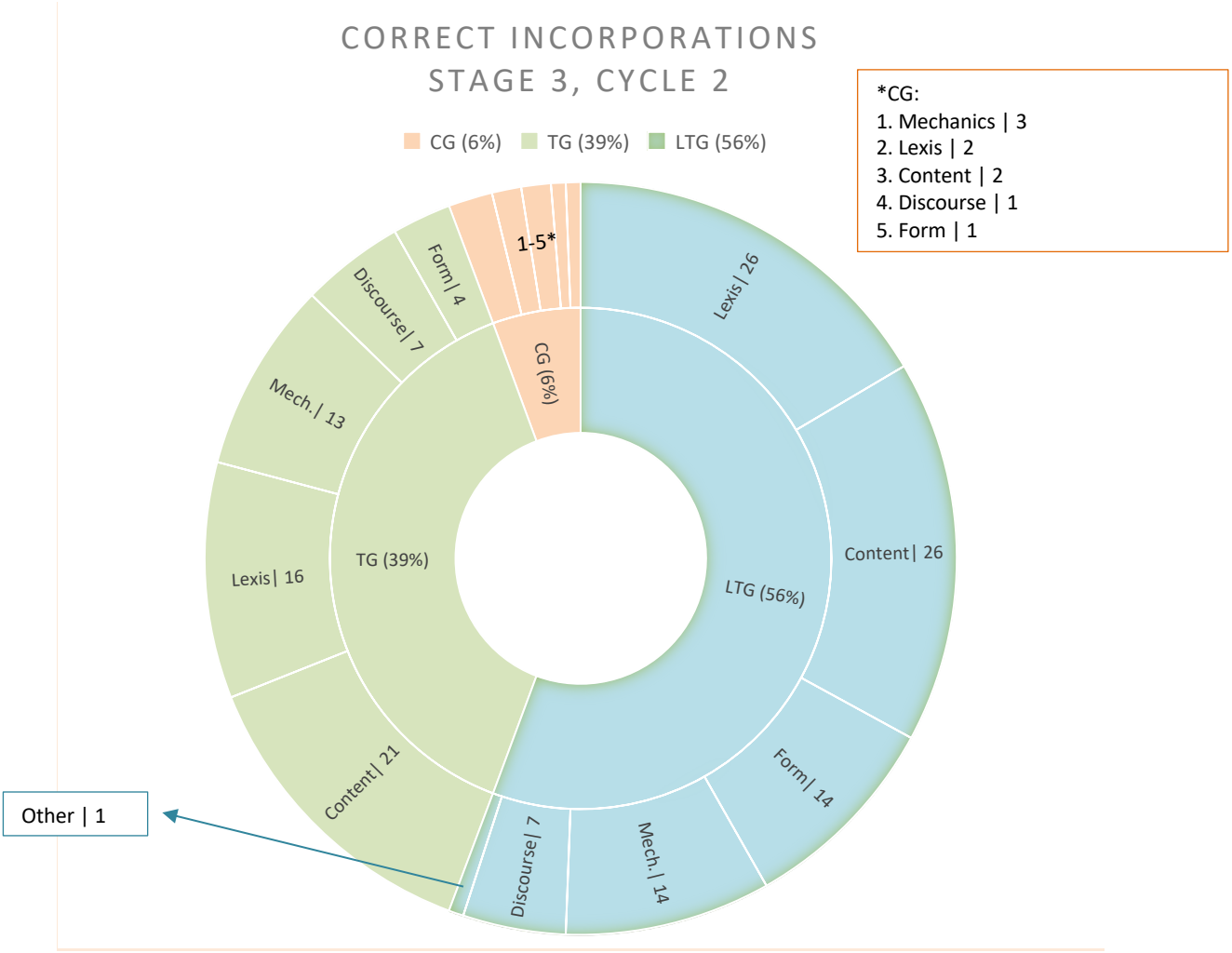
Graph 1. Correct incorporations at Stage 3 of Cycle 1



Turning now to Cycle 2, Graph 2 provides the correct features incorporated into the children's second draft. Out of a total of 182 incorporations, all dyads incorporated 158 (85%) features correctly. Six percent (6%, n = 9) of these correct incorporations belong to the self-correction group, 39% (n = 61) to the TG and 56% (n = 88) to the LTG. Unlike Cycle 1, significant differences were observed between the three groups (z = 7.04, p = < .000 between TG and CG; z = 9.63, p = 0 between CG and LTG; z = 3.04, p = .002 between TG and LTG), which points to a potential effect of a long exposure to models on the accuracy of incorporations. As for the total number of items incorporated by each group, we found that the CG incorporated 75% of the features correctly, the TG 84% of them and the 91% corresponds to the LTG.

In relation to categories, the LTG was accurate in their responses to content and lexis alike and, to a lesser extent (almost half those of lexis and content), to form and mechanics. The TG upgraded their texts mainly with respect to content but also to lexis, mechanics, discourse and form, whilst the CG's few correct incorporations had to do with mechanics followed by lexis, content, discourse and form (see Appendix 32).

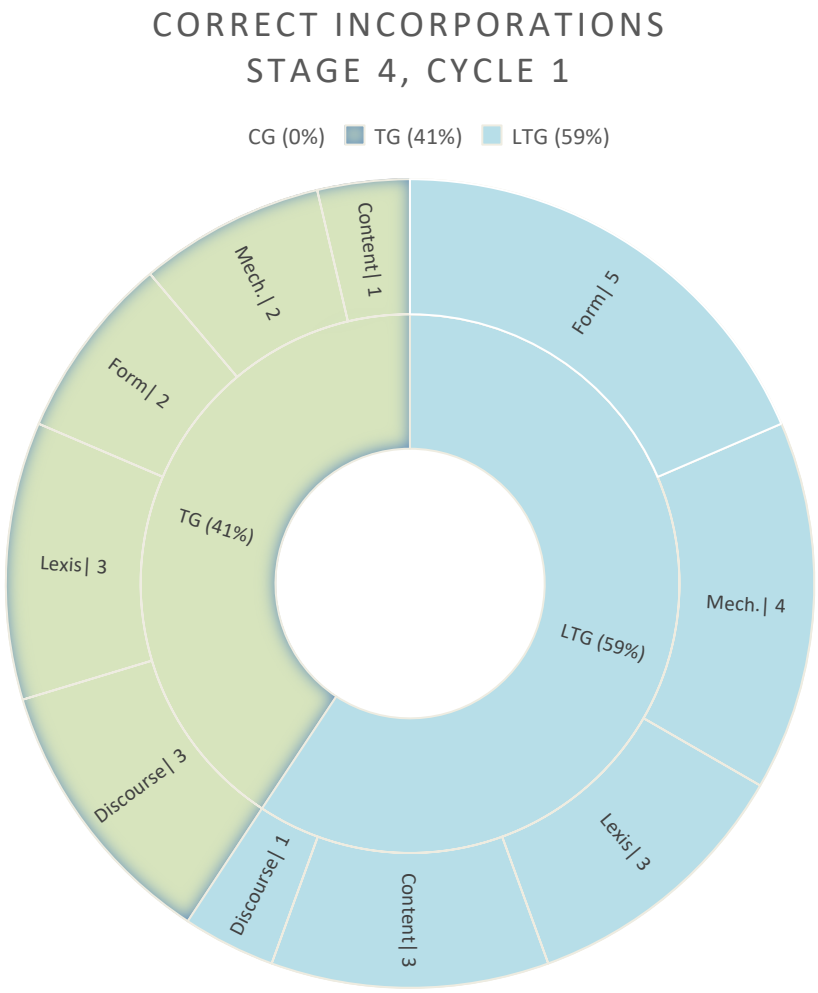
Graph 2. Correct incorporations at Stage 3 of Cycle 2



Graph 3 presents the correct revisions made at Stage 4 of Cycle 1. What stands out in this graph is that all correct answers (27 (63%) out of a total of 43 incorporations) were entirely provided by the treatment groups. As a matter of fact, let us remember that the CG did not incorporate any feature on the post-test of Cycle 1. Therefore, focusing on the treatment groups alone, at Stage 4, the TG correctly incorporated 41% (n = 11) of all acceptable revisions and the remainder was incorporated by the LTG (59%, n = 16). Although the difference expressed in percentages may seem large, chi-squared tests showed no significant differences between them (z = 1.36, p = .173). As for the percentage of acceptable incorporations out of each group's revisions, the self-correction group incorporated 0%, the TG 73% and the LTG 64% of the features correctly.

In terms of the most correctly incorporated types of features, the graph shows that a roughly even distribution across categories was achieved. In addition, similar values are observed for both groups taking into account that very few items were incorporated at Stage 4 (see Appendix 33).

Graph 3. Correct incorporations at Stage 4 of Cycle 1



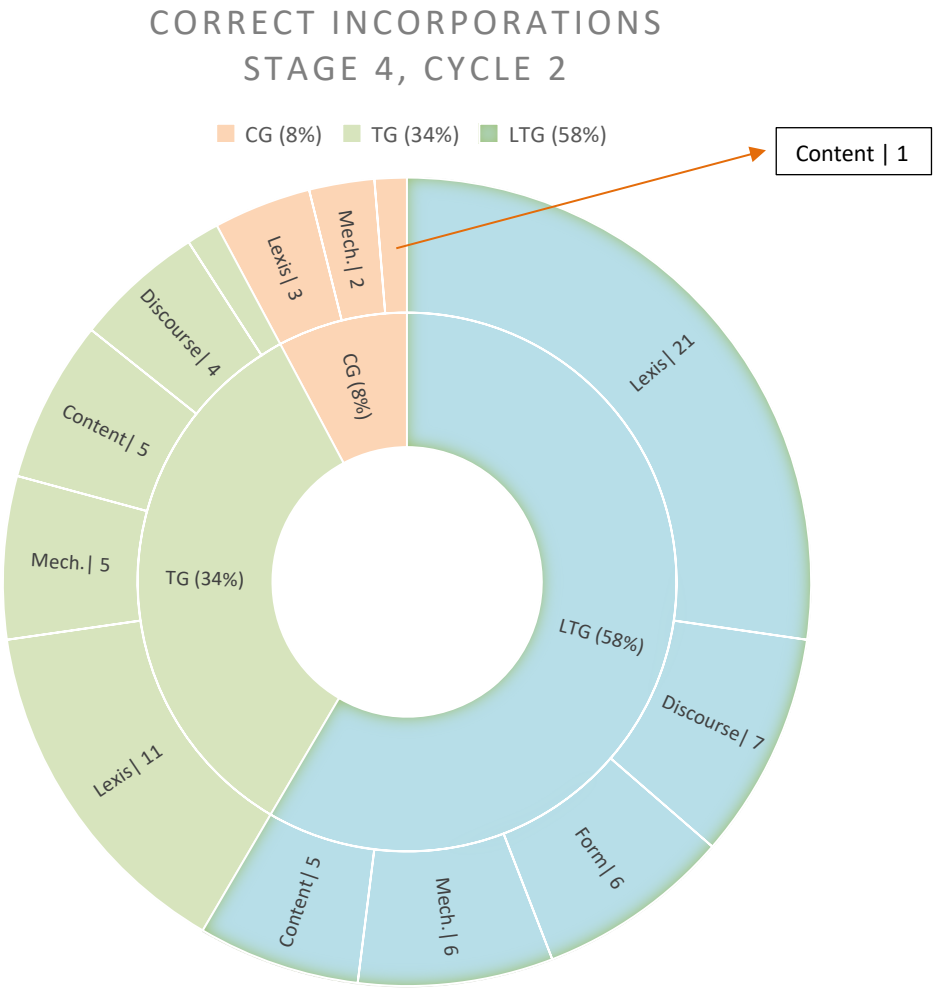
Despite the differences observed between the treatment groups and the CG at Stage 3 and 4 of Cycle 1, statistical tests did not render differences between Stage 3 and 4 in the total number of correct incorporations ($z = 2.14$, $p = .005$) nor in the frequencies of each group ($z = 1.27$, $p = .197$ for CG; $z = 1.02$, $p = .304$ for TG; $z = 2.20$, $p = .067$ for LTG). These results suggest that the children provided proportionally the same accurate answers in the rewriting stage as on the post-test, and that groups did not behave

differently in any of the stages. However, we must interpret these comparisons with caution, since we are contrasting the same procedures (rewriting) but with different prompts.

Finally, from Graph 4, which shows the accurate incorporations into the final post-test, we can see similar percentages to those observed in the previous stage. The children incorporated a total of 96 features noticed and unnoticed out of which 77 (80%) were native-like. The CG's correct incorporations constitute 8% (n = 6) of the total, the TG incorporated 34% (n = 26) of all revisions correctly, and the LTG's correct revisions represent 58% (n = 45) of the total number of correct answers. Again, significant differences were observed between the three groups (z = 3.97, p = < .000 between TG and CG; z = 6.67 p = < .000 between CG and LTG; z = 3.07, p = .002 between TG and LTG). Therefore, it seems that the impact of a long-term treatment with model texts is still visible on the post-test. Looking at the percentage of acceptable changes incorporated by each group out of the total number of incorporations, 67% corresponds to the CG, 79% to the TG, and 88% to the LTG.

As for categories, the correct revisions made by the LTG were mostly lexical, but also mechanical and formal. The TG was also accurate in their revisions made mainly to lexis but also to mechanics, discourse and content, whereas the 6 accurate answers provided by the CG were distributed among lexis, mechanics and content (see Appendix 34).

Graph 4. Correct incorporations at Stage 4 of Cycle 2



As it happened in the comparison between stages 3 and 4 of Cycle 1, no differences were observed between Stage 3 and 4 in the total number of acceptable revisions (z = 1.44, p = .147) nor in the frequencies of each group (z = 0.98, p = .325 for CG; z = 1.28, p = .198 for TG; z = 0.70, p = .483 for LTG). The data reported here appear to support the assumption that, irrespective of the feedback condition, the pairs behaved similarly in both stages of each cycle, at least with respect to frequency of correct incorporations. As far as across-cycle analyses are concerned, the difference in the total number of correct revisions was significant. More specifically, at Stage 3 of Cycle 2 the children altogether made a significantly higher number of acceptable revisions (z = 2.20, p = .027) than at the same point in Cycle 1, and that was also the case at Stage 4 of Cycle 2 (z = 2.18, p = .028). On narrowing down the analysis, we realized that only the LTG

showed an improvement in the accuracy of features incorporated into both redrafts written in Cycle 2 (Stage 3: $z = 1.90$, $p = .056$ for CG; $z = 1.67$, $p = .094$ for TG; $z = 2.84$, $p = .004$ for LTG. Stage 4: $z = 1.89$, $p = .067$ for CG; $z = 1.24$, $p = .212$ for TG; $z = 2.76$, $p = .006$ for LTG). Therefore, evidence was found here for an effect of models on the accuracy of revisions made by the children in the long term.

6.1.2.1 Summary of the main findings for research question 2

To sum up, the aim of this second research question was to determine the number and types of (acceptable) revisions made by the learner pairs to their drafts and the relationship of these incorporations to both the writing (Stage 1) and comparison (Stage 2) stages. According to the data obtained from Stages 3 and 4 of Cycle 1, we can draw several conclusions. Regarding number of incorporations, the statistical analysis broadly supports the fact that the children in the treatment groups incorporated significantly more features than the CG at both stages 3 and 4. Therefore, models proved to have a statistically significant effect on the number of revisions made by the children to their subsequent texts. However, not surprisingly, the total number of revisions made by all pairs at Stage 4 was substantially low compared to Stage 3 given that the dyads were provided with a different picture. In terms of the types of revisions made by the learner pairs to their drafts, a clear distinction is also drawn between the CG and the treatment groups in the sense that the former incorporated mostly formal features while the latter made lexical and content-related revisions to their first rewritings. Stage 4 presents a similar picture, but no significant differences were found across categories or groups. With respect to the relationship of these incorporations to both the writing and comparison stages, the data of both stages show that most of the features incorporated by the model groups either had not been previously reported or had been identified at the moment of comparison, while the majority of the changes introduced by the CG had their origin in the writing stage.

In Cycle 2 we could observe the first differences between the three groups in the number of revisions made, which suggests a positive correlation between the long-term treatment with models and the number of incorporations both in the rewriting stage and

on the post-test. In the case of the CG, no category stood out from the rest, most of their revisions had originated at the moment of writing their first draft and no FUIs were found. These results remained the same for the post-test. Interestingly, the treatment groups' most numerous incorporations into their immediate redraft coincide with the categories most attended to during the comparison stage of this same cycle. Namely, the TG incorporated mostly CREs, lexical and mechanical features and in the case of the LTG, lexis and content excelled from the rest of the categories, but a statistically considerable number of grammatical and mechanical revisions were also made at this point. At Stage 4, the TG's main incorporations were lexical and mechanical in nature, while in the case of the LTG, statistical tests showed a clear supremacy of lexis over any other category. The main difference between the treatment groups at both stages 3 and 4 lies in the fact that most of the items incorporated by the TG either had been identified at Stage 2 or had not been previously verbalized, whereas the LTG's revisions originated mainly at both the initial stage and the comparison stage with FUIs accounting for more than half of the total incorporations. Respecting inter-subject differences, the treatment groups made a higher number of content-related and lexical revisions than the CG at Stage 3. Moreover, the LTG proved to incorporate not only more vocabulary at Stage 3 and 4, but also more formal features into their second draft than their counterparts. Therefore, a long exposure to models could be a major factor, if not the only one, causing the children to incorporate more lexical and formal features than those benefitting from no treatment at all.

We also examined the individual development of each group in terms of frequency of revisions and types of categories incorporated. As for frequency of revisions, results revealed that the CG incorporated fewer features at Stage 3 of Cycle 2. While no difference was found in the total number of items incorporated for the treatment groups, statistical analyses did show differences in the types of categories. At Stage 3, the TG showed an increase in mechanical incorporations, which largely coincides not only with the rise in mechanical LREs noticed at Stage 2, but also with the increase in the incorporation of such features at Stage 3. On the other hand, significant differences in lexis were detected for the LTG only at Stage 4.

Finally, apart from analyzing the types and frequency of revisions, we aimed to explore how many of those features were incorporated correctly and whether or not there was an

effect of models on accuracy in the short and long run. Firstly, we found a similar high percentage of total number of correct answers provided by the three groups together throughout the four rewriting tasks of both cycles, with a slight increase in cycle 2. In fact, this rise was found to be statistically significant for both Stage 3 and 4 of Cycle 2 with respect to Cycle 1. Upon zooming in on group differences, we observed that it was the LTG that triggered this difference, since it was the only group showing an improvement in the accuracy of features incorporated into their second writing after five months of treatment. Secondly, within stages, the treatment groups always seemed to make more native-like revisions than the self-correction group in Cycle 1, whereas in Cycle 2, the LTG managed to surpass the other two groups. Therefore, it seems that the impact of a long-term treatment with model texts is still visible on the post-test. Thirdly, across stages, no differences were observed between Stage 3 and 4 within each cycle. Last, but not least, in general terms the vast majority of the upgraded revisions made by the treatment groups were content- and lexis-related followed by form, mechanics, discourse and other issues, whereas most of the accurate answers provided by the CG were mostly formal and mechanical in nature.

6.1.3 Trajectories

This first module has focused on the children' oral production through the analysis of CREs and LREs as well as on the connection between the oral and written modes. Now that we know the details, we considered it important to have a quick look at the big picture. Although we have already commented thoroughly upon the data provided below, there is some information that has not been considered thus far and will be discussed here as an add-on. The following six tables present an overview of the tracking of the episodes noticed and the revisions made by the groups throughout the four stages of both cycles. In Table 43, which illustrates the data for the CG in Cycle 1, we can see that the self-correction group verbalized 154 PFNs at Stage 1 and 51 FNs at Stage 2. Out of these 51 FNs, only 9 (18%) had been PFNs. In their subsequent texts, the children incorporated 19 features from stages 1 and 2 at Stage 3, and 3 at Stage 4. Table 44 presents a very similar case. Eight (15%) of the features noticed at Stage 2 had been PFNs at Stage 1. Therefore, the group behaved similarly in both stages as far as proportion of PFNs over FNs is concerned. With relation to incorporations, at Stage 3, the CG incorporated 19 (37%) features, and at Stage 4 they incorporated only 3 (6%)

items out of a total of 51 noticed forms in Cycle 1. In Cycle 2, the children made 12 (23%) changes at Stage 3, and 9 (23%) changes on the delayed post-test out of 52 FNs.

Table 43. Tracking of the episodes noticed by the CG across stages and into their revised texts in Cycle 1

	CYCLE 1	**CREs**		**LREs**										
				MECHANICS		**LEXIS**		**FORM**		**DISCOURSE**		**OTHER**		**TOTAL**
		FO	%	FO	%	FO	%	FO	%	FO	%	FO	%	FO
CG	**PFNs** (S1)	12	8	42	27	46	30	46	30	6	4	2	1	**154**
	FNs (S2)	2	4	12	23	6	12	22	43	9	18	0	0	**51**
	It was a PFN	0	0^1 | 0^2	2	17 | 22	4	67 | 44	3	14 | 33	0	0 | 0	0	0 | 0	**9**
	FNI (S3)	1	5	4	21	4	21	10	53	0	0	0	0	**19**
	FUI	0	0	0	0	0	0	0	0	0	0	0	0	**0**
	FNI (S4)	0	0	1	33	0	0	2	67	0	0	0	0	**3**
	FUI	0	0	0	0	0	0	0	0	0	0	0	0	**0**

Percentage calculated out of the total number of [1]FNs, [2]PFNs

Table 44. Tracking of the episodes noticed by the CG across stages and into their revised texts in Cycle 2

	CYCLE 2	**CREs**		**LREs**										
				MECHANICS		**LEXIS**		**FORM**		**DISCOURSE**		**OTHER**		**TOTAL**
		FO	%	FO	%	FO	%	FO	%	FO	%	FO	%	FO
CG	**PFNs** (S1)	6	4	54	40	28	21	40	30	3	2	4	3	**135**
	FNs (S2)	4	8	20	38	5	10	14	27	6	11	3	6	**52**
	It was a PFN	1	25 | 12	1	5 | 12	2	40 | 25	4	28 | 50	0	0 | 0	0	0 | 0	**8**
	FNI (S3)	2	17	5	42	2	17	2	17	1	8	0	0	**12**
	FUI	0	0	0	0	0	0	0	0	0	0	0	0	**0**
	FNI (S4)	1	11	3	3	4	44	1	11	0	0	0	0	**9**
	FUI	0	0	0	0	0	0	0	0	0	0	0	0	**0**

When it comes to the treatment groups, apart from the information concerning whether or not the features noticed by the children had initially been PFNs, we can also get to know how many PFNs were solvable with the model. Tables 45 and 46 display the traceability of the episodes generated by the TG in Cycle 1 and 2, respectively. In cycle 1, out of a total of 186 PFNs, 132 (71%) of them could be solved with the help of the model text. The pairs noticed, however, 118 features at Stage 2 out of which 29 were PFNs, and 34 (29%) were incorporated at the rewriting stage and only 7 (6%) on the delayed post-test. Therefore, the children in the TG condition observed 29 features out of 132 solvable ones. That makes 22% of all solutions to their initial difficulties. As for categories, where the children proportionally found most answers to their questions were in content (40%) and lexis (32%) over discourse (25%), mechanics (7%) and form (9%).

In Cycle 2, the dyads experienced problems with 137 features at the moment of writing their first draft. They could find solution to 74 (54%) of those problems at the comparison stage, but they only noticed 16 solutions. Again, that makes 22% of all the solvable features. Concerning categories, the children in the TG could solve 32% of their difficulties with vocabulary, 27% of their issues with content, 22% of their problems with mechanics (let us remember that this group showed a statistically significant increase in mechanical LREs in Cycle 2), 14% of their formal issues and 11% of their discursive ones. It can therefore be assumed that models seem to offer solutions mostly concerned with content and lexis. With respect to incorporations, the participants incorporated 40% of the noticed forms into their second draft and 18% of 111 FNs into the text written at Stage 4.

Table 45. Tracking of the episodes noticed by the TG across stages and into their revised texts in Cycle 1

	CYCLE 1	**CREs**		**LREs**													
				MECHANICS		**LEXIS**		**FORM**		**DISCOURSE**		**OTHER**		**TOTAL**			
		FO	%	FO	%	FO	%	FO	%	FO	%	FO	%	FO			
	PFNs (S1)	27	14	37	20	63	34	49	26	9	5	1	0	**186**			
	Solvable	20	74[1] | 15[2]	15	40 | 11	47	75 | 36	46	94 | 35	4	44 | 3	0	0 | 0	**132**		
	FNs (S2)	42	35	4	3	37	31	12	10	12	10	11	9	**118**			
TG	**It was a PFN**	8	19[3] | 40[4]	1	25 | 7	15	40 | 32	4	33 | 9	1	8 | 25	0	0 | 0	**29**		
	FNI (S3)	6	18	2	6	14	42	8	24	3	9	1	3	**34**			
	FUI	12	36	5	15	12	36	4	12	0	0	0	0	**33**			
	FNI (S4)	0	0	0	0	4	57	1	14	2	28	0	0	**7**			
	FUI	1	12	2	25	3	37	1	12	1	12	0	0	**8**			

Percentage calculated out of the total number of [1]PFNs, [2]solvable features, [3]FNs, [4]solvable features

Table 46. Tracking of the episodes noticed by the TG across stages and into their revised texts in Cycle 2

	CYCLE 2	**CREs**		**LREs**													
				MECHANICS		**LEXIS**		**FORM**		**DISCOURSE**		**OTHER**		**TOTAL**			
		FO	%	FO	%	FO	%	FO	%	FO	%	FO	%	FO			
TG	**PFNs** (S1)	16	12	35	25	37	27	36	26	11	8	2	1	**137**			
	Solvable	15	94 | 20	9	26 | 12	19	51 | 26	21	58 | 28	9	82 | 12	1	50 | 1	**74**		
	FNs (S2)	29	26	25	22	27	24	14	13	6	6	10	9	**111**			
	It was a PFN	4	14 | 27	2	8 | 22	6	22 | 32	3	21 | 14	1	17 | 11	0	0 | 0	**16**		
	FNI (S3)	11	23	13	27	14	29	6	12	4	8	0	0	**48**			
	FUI	11	44	4	16	7	28	0	0	3	12	0	0	**25**			
	FNI (S4)	3	15	6	30	6	30	2	10	3	15	0	0	**20**			
	FUI	3	23	2	15	7	54	0	0	1	8	0	0	**13**			

If we now explore the trajectory of the LTG, Table 47 features that the group identified 173 content and linguistic problems at Stage 1 of Cycle 1, 109 (63%) of which were solvable with the model. At the moment of comparison, they noticed 108 features in the model, of which 50 (46%) and 14 (13%) were incorporated into their revisions at Stage 3 and 4, respectively. However, only 30 (28%) of them had originally been PFNs. Consequently, the pairs in the long-term treatment condition found solution to 28% of their initial problems, a result which is rather consistent with that of the TG. Taking a closer look at the different categories, we can see that the children were able to find solutions mostly related to lexis (63%), discourse (37%) and content (20%) and, to a lesser extent, mechanics (7%) and form (7%). Although the number of discursive features is very low in all their occurrences, it is important to underscore here the considerable attention that this category has been given at the second stage of this first cycle, even more than content. Nevertheless, these results are not surprising, since they match those observed when we analyzed what this group focused on while dealing with the model, namely, discursive elements were highly attended to.

Moving on to Cycle 2, out of 88 PFNs, 50 were solvable with the model. Put it differently, more than half (57%) of the problems the children experienced when writing the text had a solution in the model. When the pairs were provided with the model, they noticed 71 features, of which 12 had been PFNs. Thus, 24% of the solutions offered by the native text were detected by the children in the LTG. This last percentage largely coincides with that of the previous cycle (28%) and also of the TG (22% for Cycle 1, 22% for Cycle 2). As a matter of fact, we analyzed these relationships statistically, and no differences were found between the treatment groups ($z = 0.99$, $p = .318$ for Cycle 1; $z = 0.31$, $p = .756$ for Cycle 2). As regards categories, the three problems the children experienced with content-related issues at Stage 1 were clarified with the model (100%) at Stage 2, 45% of their lexical PFNs were solved, 25% in the case of mechanics and 13% in the case of form. In consequence, content and lexis remain in the lead. Concerning incorporations, at the rewriting stage the children incorporated 58% of the noticed units and on the delayed post-test they used 34% of the FNs at Stage 2. Table 48 details these results.

Table 47. Tracking of the episodes noticed by the LTG across stages and into their revised texts in Cycle 1

	CYCLE 1	**CREs**		**LREs**												
				MECHANICS		**LEXIS**		**FORM**		**DISCOURSE**		**OTHER**		**TOTAL**		
		FO	%	FO	%	FO	%	FO	%	FO	%	FO	%	FO		
LTG	**PFNs** (S1)	17	10	47	27	40	23	57	33	11	6	1	0	**173**		
	Solvable	15	88 | 14	14	30 | 13	30	75 | 27	42	74 | 38	8	73 | 7	0	0 | 0	**109**	
	FNs (S2)	37	34	5	5	38	35	15	14	8	7	5	5	**108**		
	It was a PFN	3	8 | 20	1	20 | 7	19	50 | 63	3	20 | 7	3	37 | 37	0	20 | 0	**30**	
	FNI (S3)	13	26	5	10	15	30	9	18	8	16	0	0	**50**		
	FUI	6	23	4	15	12	46	3	11	1	4	0	0	**26**		
	FNI (S4)	3	21	0	0	2	14	6	43	3	21	0	0	**14**		
	FUI	1	9	5	45	4	36	0	0	1	9	0	0	**11**		

Table 48. Tracking of the episodes noticed by the LTG across stages and into their revised texts in Cycle 2

	CYCLE 2	**CREs**		**LREs**												
				MECHANICS		**LEXIS**		**FORM**		**DISCOURSE**		**OTHER**		**TOTAL**		
		FO	%	FO	%	FO	%	FO	%	FO	%	FO	%	FO		
LTG	**PFNs** (S1)	3	3	21	24	22	25	26	29	14	16	2	2	**88**		
	Solvable	3	100 | 6	8	38 | 16	11	50 | 22	15	58 | 30	12	86 | 24	1	50 | 2	**50**		
	FNs (S2)	18	25	13	18	13	18	16	22	3	4	8	11	**71**		
	It was a PFN	3	17 | 100	2	15 | 25	5	38 | 45	2	12 | 13	0	0 | 0	0	0 | 0	**12**		
	FNI (S3)	9	22	10	24	13	32	7	17	2	5	0	0	**41**		
	FUI	17	30	6	11	20	36	7	12	5	9	1	1	**56**		
	FNI (S4)	2	8	5	21	9	37	6	25	2	8	0	0	**24**		
	FUI	3	11	3	11	16	59	0	0	5	18	0	0	**27**		

6.2. Module 2: The effects of models on the children's written production

This second module encompasses research questions three and four and revolves around the impact of model texts on the children's written production. In the following section, we will provide the results obtained for the third research question. In order to answer it, the models of the children's final drafts were assessed using type of clause, CAF measures and a holistic rubric. The findings about the fourth research question, which deals with the use of textual enhancement of third person possessives, will be presented afterwards.

6.2.1 Results for research question 3

The third research question is stated here again for the reader's convenience:

3. Do models help improve the written production of primary EFL students in the short and long run?

To examine the impact of the models on the children's written production in the short-term, the quality of the third draft (Stage 4) in relation to the first one (Stage 1) was measured in both cycles (draft 1 vs. draft 3 and draft 4 vs. draft 6). In order to examine the effect of the feedback in the long run, the first and last draft (draft 1 vs. draft 6) were compared. Starting with the results obtained for type of clause and CAF, the rates for the CG, the TG and the LTG are presented in the following three tables, respectively. Table 49 below contains descriptive statistics of type of clause and CAF measures altogether per group and draft.

Table 49. Descriptive statistics on frequency of occurrences, means and standard deviations of type of clause and CAF measures per group and draft

		CG			TG			LTG	
		FO	*Mean*	*SD*	*FO*	*Mean*	*SD*	*FO*	*Mean*	*SD*
1[1]	*Pre-clause*	16	1.78	1.72	16	1.45	0.93	13	1.30	.95
	Proto-clause	36	4.00	1.41	48	4.36	1.69	49	4.90	1.37
	Clause	40	4.44	1.88	56	5.09	3.02	51	5.10	2.23
	Sub.clause	8	0.89	1.05	14	1.27	1.27	14	1.40	1.07
	Coord.clause	31	3.44	1.67	44	4.00	2.32	36	3.60	1.84
	Lexical diversity	42	4.66	0.87	53	4.85	0.30	47	4.71	0.58
	Accuracy	38	4.17	1.61	37	3.34	1.11	30	3.01	0.35
	Fluency	469	52.11	10.66	686	62.36	10.52	710	71.00	16.75
3[2]	*Pre-clause*	12	1.33	1.22	2	0.18	0.60	0	0.00	0.00
	Proto-clause	25	2.78	1.79	40	3.64	2.29	47	4.70	1.49
	Clause	50	5.56	1.33	72	6.55	2.66	57	5.70	1.70
	Sub.clause	16	1.78	2.05	31	2.82	1.94	20	2.00	1.41
	Coord.clause	39	4.33	1.80	42	3.82	1.94	41	4.10	2.13
	Lexical diversity	40	4.45	0.60	52	4.76	0.66	45	4.51	0.39
	Accuracy	37	4.15	1.83	33	2.99	1.16	34	3.37	1.08
	Fluency	510	56.67	12.27	694	63.09	16.06	607	60.70	13.11
4[3]	*Pre-clause*	6	0.67	1.32	7	0.45	0.69	1	0.10	.32
	Proto-clause	39	4.33	1.58	40	3.64	1.50	34	3.40	1.90
	Clause	24	2.67	2.83	34	3.09	2.07	37	3.70	1.49
	Sub.clause	9	1.00	1.00	9	0.82	0.98	5	0.50	.53
	Coord.clause	27	3.00	1.00	23	2.09	0.94	21	2.10	1.10
	Lexical diversity	37	4.14	0.55	48	4.36	0.67	37	3.66	0.50
	Accuracy	39	4.39	1.63	40	3.64	1.21	24	2.35	0.99
	Fluency	366	40.67	10.55	521	47.36	13.76	450	45.00	10.68
6[4]	*Pre-clause*	5	0.56	0.88	1	0.09	0.30	1	0.10	.32
	Proto-clause	32	3.56	1.42	32	2.91	1.38	21	2.10	2.13
	Clause	22	2.44	1.51	38	3.45	1.92	57	5.70	1.49
	Sub.clause	7	0.78	0.83	15	1.36	1.29	3	0.30	.67
	Coord.clause	16	1.78	1.30	15	1.36	0.50	24	2.40	1.43
	Lexical diversity	36	3.99	0.53	45	4.11	0.40	40	3.99	0.50
	Accuracy	38	4.27	1.87	37	3.37	1.23	19	1.94	1.12
	Fluency	348	38.67	9.29	464	42.18	9.70	465	46.50	7.62

[1]*Draft 1,* [2]*draft 3,* [3]*draft 4,* [4]*draft 6*

In what follows, the statistical results of each aspect will be shown. As for pre-clauses, within the global domain of type of clause, we used the Greenhouse-Geisser correction

since the sphericity assumption was not met (X^2 = 13.74; p = .017). Table 50 provides the results of the mixed ANOVA for the pre-clause parameter. From the table we can see that there is a main effect for Pre-clause ($F_{(2.21,59.72)}$ = 16.92; p = < .000; η_p^2 = 0.38), Group ($F_{(2,27)}$ = 3.74; p = .037; η_p^2 = 0.21), and for the Pre-clause x Group ($F_{(4.42,59.72)}$ = 8.03; p = .048; η_p^2 = 0.07) interaction.

Table 50. Mixed-model ANOVA results for pre-clause

	Stage 2				
	***df*1**	***df*2**	***F***	***p***	η_p^2
Pre-clause	2.21	59.72	16.92	< .000	0.38
Pre-clause * Group	4.42	59.72	8.03	.048	0.07
Inter-subjects					
Group	2	27	3.74	.037	0.21

Pairwise comparisons with Bonferroni adjustment showed some statistically significant intra- and inter-subject differences. As for the former, the TG significantly reduced the number of pre-clauses from drafts 1 to 3 (p = .004, 95% CI [0.33, 2.21], d = 0.70), 4 to 6 (p = .045, 95% CI [-0.40, 1.13], d = 0.25) and 1 to 6 (p = .024, 95% CI [0.13, 2.60], d = 0.58). The LTG also improved significantly from their first to their third composition (p = .004, 95% CI [0.31, 2.29], d = 0.68), and from the first to the last one (p = .031, 95% CI [-0.25, 3.20], d = 0.48). The self-correction group showed no improvement in terms of pre-clauses. Group differences were only observed on the first delayed post-test (draft 3) where both treatment groups outperformed the CG (p = .007, 95% CI [0.28, 2.02], d = 0.61 for CG-TG; p = .002, 95% CI [0.44, 2.23], d = 0.70 for CG-LTG).

Regarding proto-clauses, Mauchly's test did not indicate any violation of sphericity (X^2 = 4.07; p = .539). Table 51 displays the results of the mixed ANOVA for the proto-clause parameter. As can be seen, a main effect was found for Proto-clause ($F_{(3,81)}$ = 4.73; p = .004; η_p^2 = 0.14) and for the interaction between Proto-clause and Group ($F_{(6,81)}$ = 2.82; p = .044; η_p^2 = 0.14).

Table 51. Mixed-model ANOVA results for proto-clause

	Stage 2				
	***df*1**	***df*2**	***F***	***p***	η_p^2
Proto-clause	3	81	4.73	.004	0.14
Proto-clause * Group	6	81	2.82	.044	0.14
Inter-subjects					
Group	2	27	0.05	.943	0

Post-hoc tests again revealed statistical differences between draft 1 and draft 6 for both model groups (p = .032, 95% CI [-1.61, 3.52], d = 0.37 for TG; p = .006, 95% CI [0.63, 4.97], d = 0.67 for LTG), but not for the CG. This time, no between-group differences were observed for this aspect.

Concerning clauses, sphericity was met (X^2 = 8.34; p = .139), as indicated by Mauchly's test. Table 52 presents the results of the mixed ANOVA for the clause parameter. There is a significant main effect for Clause ($F_{(3,81)}$ = 14.18; p = < .000; η_p^2 = 0.34), for Group ($F_{(2,27)}$ = 5.52; p = .043; η_p^2 = 0.12), and for the Clause x Group ($F_{(6,81)}$ = 7.96; p = .027; η_p^2 = 0.46) interaction.

Table 52. Mixed-model ANOVA results for clause

	Stage 2				
	***df*1**	***df*2**	***F***	***p***	η_p^2
Clause	3	81	14.18	.000	0.34
Clause * Group	6	81	7.96	.027	0.46
Inter-subjects					
Group	2	27	5.52	.043	0.12

Further pairwise comparisons revealed statistically significant within-group and between-group differences. In this case, we found that the CG wrote statistically fewer clauses in draft 6 in contrast to draft 1 (p = .002, 95% CI [-0.48, 5.30], d = 0.41), and the LTG produced a significantly higher number of clauses in draft 6 in comparison with draft 4 (p = .012, 95% CI [-3.66, -0.34], d = -0.63). No differences across drafts were observed for the TG. Looking at between-group differences, post-hoc analyses revealed that the LTG outperformed both the CG (p = .001, 95% CI [1.30, 5.21], d = 0.78) and the TG (p = .014, 95% CI [0.39, 4.11], d = 0.56) in draft 6.

Moving on to the area of complexity, and more specifically to subordinate clauses, we used the Greenhouse-Geisser correction since the sphericity assumption was rejected (X^2 = 20.13; p = .001). Table 53 provides the statistics for subordinate clauses. It can be seen from the data that there is a main effect for Subordinate clause ($F_{(2.07,56.10)}$ = 11.49; p = < .000; η_p^2 = 0.29), and for the Subordinate clause by Group ($F_{(4.15,56.10)}$ = 9.62; p = .037; η_p^2 = 0.15) interaction effect.

Table 53. Mixed-model ANOVA results for subordinate clauses

	Stage 2				
	***df*1**	***df*2**	***F***	***p***	η_p^2
Subordinate clause	2.07	56.10	11.49	<.000	0.29
Subordinate clause * Group	4.15	56.10	9.62	.037	0.15
Inter-subjects					
Group	2	27	1.15	.329	0.07

Further post-hoc tests located these differences for both experimental groups between drafts 1 and 3 (p = .006, 95% CI [0.45, 3.55], d = -0.67 for TG; p = .024, 95% CI [-2.29, 1.09], d = -0.18 for LTG) and 4 and 6 (p = .048, 95% CI [0.62, 3.16], d = 0.37 for TG; p = .001, 95% CI [1.34, 3.22], d = 0.13 for LTG). No statistically significant differences were found for the CG or between groups.

With reference to coordinate clauses, the assumption of sphericity was not rejected (X^2 = 10.93; p = .053). The results obtained from the analysis of coordinate clauses are set out in Table 54. Only a significant main effect for Coordinate clause ($F_{(3,81)}$ = 18.11; p = < .000; η_p^2 = 0.40) was found. As our omnibus ANOVA is not statistical in the variables we are interested in, we will not continue with post-hocs.

Table 54. Mixed-model ANOVA results for coordinate clauses

	Stage 2				
	***df*1**	***df*2**	***F***	***p***	η_p^2
Coordinate clause	3	81	18.11	<.000	0.40
Coordinate clause * Group	6	81	1.05	.395	0.07
Inter-subjects					
Group	2	27	0.24	.787	0.01

The next piece of the output, shown in Table 55, contains the results for lexical diversity. Mauchly's tests indicated that sphericity was not violated (X^2 = 6.27; p = .281). The table below shows that there is a significant main effect for Lexical diversity ($F_{(3,81)}$ = 23.18; p = < .000; η_p^2 = 0.46), Group ($F_{(2,27)}$ = 9.33; p = .028; η_p^2 = 0.09), and for the interaction between Lexical diversity x Group ($F_{(6,81)}$ = 16.43; p = .021; η_p^2 = 0.17).

Table 55. Mixed-model ANOVA results for lexical diversity

	Stage 2				
	***df*1**	***df*2**	***F***	***p***	η_p^2
Lexical diversity	3	81	23.18	<.000	0.46
Lexical diversity * Group	6	81	16.43	.021	0.17
Inter-subjects					
Group	2	27	9.33	.028	0.09

For the Bonferroni post-hoc test, there is a statistical difference between drafts 1 and 6 within the three groups. That is, the lexical repertoire of the children in both the CG (p = .013, 95% CI [10.53, 123.03], d = 0.62) and the TG (p = .002, 95% CI [23.21, 124.97], d = 0.76) appeared to be significantly richer in draft 1 as opposed to draft 6. On the contrary, the LTG (p = .004, 95% CI [18.84, 125.56], d = 0.70) seemed to increase the number of lexical words to the extent of this rise being statistically significant with respect to their very first composition. For the variable of lexical diversity, the omnibus ANOVA also showed a statistical difference between groups. Post-hocs located these differences in drafts 4 and 6 between the LTG and the TG (p = .034, 95% CI [-135.49, -4.28], d = -0.50 for draft 4; p = .024, 95% CI [-122.90, -6.19], d = -0.57 for draft 6) and between the LTG and the CG (p = < .000, 95% CI [-135.62, -4.67], d = -0.32 for draft 4; p = .002, 95% CI [-142,20, -8.41], d = -0.11 for draft 6), which means that the long-term group used a wider range of vocabulary in their texts than their counterparts in Cycle 2.

The output with the main results for accuracy are shown in Table 56. Sphericity was met (X^2 = 3.56; p = .615) and, in this case, we observe a main effect for Group ($F_{(2,27)}$ = 4.72; p = .027; η_p^2 = 0.25) and for the interaction between Accuracy and Group ($F_{(6,81)}$ = 3.51; p = .004; η_p^2 = 0.2).

Table 56. Mixed-model ANOVA results for accuracy

	Stage 2				
	***df*1**	***df*2**	***F***	***p***	η_p^2
Accuracy	3	81	1.08	.360	0.03
Accuracy * Group	6	81	3.51	.004	0.2
Inter-subjects					
Group	2	27	4.72	.017	0.25

Comparisons only found a statistical difference between drafts 1 and 6 for the LTG (p = .013, 95% CI [17.44, 196.56], d = 0.62), and a difference in drafts 4 and 6 between the LTG and the CG (p = .006, 95% CI [-355.36, -52.18], d = -0.63 for draft 4; p = .004, 95% CI [-400.17, -66.23], d = -0.65 for draft 6) and between the LTG and the TG (p = .045, 95% CI [-367.31, -43.20], d = 0.16 for draft 4; p = .038, 95% CI [-287.89, -24.76], d = 0.31 for draft 6).

The last piece of the output, presented in Table 57, provides the results of the mixed ANOVA for the area of fluency, whose sphericity was not rejected (X^2 = 7.10; p = .214). As we can see, a significant main effect was found for Fluency ($F_{(3,81)}$ = 29.17; p = < .000; η_p^2 = 0.51) and for the interaction between Fluency and Group ($F_{(6,81)}$ = 36.70; p = .002; η_p^2 = 0.32).

Table 57. Mixed-model ANOVA results for fluency

	Stage 2				
	***df*1**	***df*2**	***F***	***p***	η_p^2
Fluency	3	81	29.17	<.000	0.51
Fluency * Group	6	81	36.70	.002	0.32
Inter-subjects					
Group	2	27	2.96	.069	0.18

Further pairwise comparisons revealed that the treatment groups wrote significantly fewer words on the second post-test (draft 6) than in their very first draft (p = < .000, 95% CI [7.79, 32.57], d = 0.85 for TG; p = < .000, 95% CI [11.50, 37.50], d = 0.18 for LTG). The children in the self-correction group also showed a decline near significance with a large effect size (p = .057, 95% CI [-0.25, 27.14], d = 0.51). In other words, in overall terms, the three groups behave alike as far as number of words is concerned. No significant difference between the three groups was evident.

The following three tables offer a visual summary of the significant values obtained for each group in each of the dimensions evaluated throughout the stages and cycles. Significance is marked with a tick, while a cross is used for non-significant values.

If we zoom in on Table 58, which shows the development of the CG across drafts, we can observe that the only statistically significant result concerning type of clause can be found in the long-term, with the dyads in the self-correction group producing significantly fewer clauses in the last composition as compared to the first one. Likewise, lexical diversity and fluency underwent a statistically significant decrease from draft 1 to draft 6. On the other hand, grammatical complexity and accuracy results remained quite stable across drafts.

Table 58. Type of clause and CAF measures: summary of the CG

Dimension	**Aspect**	**CG**					
		Draft 1	**Draft 3**	**Draft 4**	**Draft 6**	**Draft 1**	**Draft 6**
Type of clause	Pre-clause	x		x		x	
	Proto-clause	x		x		x	
	Clause	x		x		√	
Complexity	Subordinate clauses	x		x		x	
	Coordinate clauses	x		x		x	
	Lexical diversity	x		x		√	
Accuracy	Errors per total number of words	x		x		x	
Fluency	Total number of words	x		x		√	

Table 59 provides the visual summary for the TG. In all three comparisons, the children seemed to produce significantly fewer pre-clauses, and also fewer proto-clauses when draft 1 and draft 6 are contrasted. The pairs also produced significantly more subordinate clauses on the post-test of both cycles (draft 3 and draft 6). Contrary to expectations, however, and in line with the CG, lexical diversity turned out to be significantly lower and pairs also wrote significantly shorter texts in the very last written task as opposed to the first one. No accuracy comparison yielded significant

differences across compositions. Therefore, models seem to have had an overall positive effect on the types of clauses and on grammatical complexity.

Table 59. Type of clause and CAF measures: summary of the TG

Dimension	Aspect	TG					
		Draft 1	Draft 3	Draft 4	Draft 6	Draft 1	Draft 6
Type of clause	Pre-clause	√		√		√	
	Proto-clause	x		x		√	
	Clause	x		x		x	
Complexity	Subordinate clauses	√		√		x	
	Coordinate clauses	x		x		x	
	Lexical diversity	x		x		√	
Accuracy	Errors per total number of words	x		x		x	
Fluency	Total number of words	x		x		√	

Table 60 presents the information about the LTG. With respect to type of clause, we found significantly fewer pre-clauses from draft 1 to draft 3 and to draft 6, but also fewer proto-clauses from draft 1 to draft 6 and more clauses from 4 to 6. As far as grammatical complexity is concerned, we can observe the same behavior as the TG, namely, significantly more subordinate clauses were found on the post-tests as compared with the children's corresponding first texts. When it comes to lexical diversity, we found a difference with respect to the other two groups, as the means do not decrease, but significantly increase when draft 1 and draft 6 are compared. Also unlike their counterparts, which showed no statistically significant changes across compositions, the pairs in the LTG made significantly fewer mistakes from draft 1 to draft 6. Finally, regarding fluency, this group behaved in the same way as the CG and the TG, as the children in the LTG wrote significantly shorter lasts texts as compared to the very first ones. As a result, we could surmise that a long exposure to model texts helped learners enhance the overall quality of their texts in terms of type of clause, grammatical complexity, lexical diversity and accuracy, but no effect was found on fluency.

Table 60. Type of clause and CAF measures: summary of the LTG

Dimension	Aspect	LTG					
		Draft 1	Draft 3	Draft 4	Draft 6	Draft 1	Draft 6
Type of clause	Pre-clause	√		x		√	
	Proto-clause	x		x		√	
	Clause	x		√		x	
Complexity	Subordinate clauses	√		√		x	
	Coordinate clauses	x		x		x	
	Lexical diversity	x		x		√	
Accuracy	Errors per total number of words	x		x		√	
Fluency	Total number of words	x		x		√	

Finally, some of the differences mentioned above become more visible when the three groups are contrasted. Table 61 displays visual information about the differences between groups across drafts. To start with, no differences were observed in draft 1 for any of the measures, which suggests that the three groups' texts were similar regarding type of clause, complexity, accuracy and fluency. Nevertheless, as the children wrote more drafts, some differences emerged. To start with, the treatment groups produced significantly fewer pre-clauses than the CG on the first post-test (draft 3) and, what is more, the pairs in the LTG wrote a significantly higher number of clauses than their counterparts on the last post-test (draft 6). On the other hand, one of the most remarkable results comes with the analysis of lexical diversity and accuracy. When the children in the LTG wrote their first texts after the four-month exposure to models, they used a significantly higher number of different words in their texts and also made significantly fewer errors than the CG and the TG, and these differences remained in draft 6. As for proto-clauses, grammatical complexity and fluency, although we found within-group differences, none of the analyses carried out here reached significance, which indicates that the different learner pairs assigned to either of the three feedback types were similar insofar as these three aspects.

Table 61. Type of clause and CAF measures: summary of differences between groups

Dimension	Aspect	Draft	Between-group comparison
Type of clause	Pre-clause	Draft 3	CG-**TG**: √
			TG-LTG: x
			CG-**LTG**: √
	Clause	Draft 6	CG-TG: x
			TG-**LTG**: √
			CG-**LTG**: √
Complexity	Lexical diversity	Draft 4	CG-TG: x
			TG-**LTG**: √
			CG-**LTG**: √
		Draft 6	CG-TG: x
			TG-**LTG**: √
			CG-**LTG**: √
Accuracy	Errors per total number of words	Draft 4	CG-TG: x
			TG-**LTG**: √
			CG-**LTG**: √
		Draft 6	CG-TG: x
			TG-**LTG**: √
			CG-**LTG**: √

Having analyzed the quantitative results, the final section of this research question will address the qualitative evaluation of the written texts. Table 62 gives us the means and standard deviations of holistic measures per group and draft.

Table 62. Descriptive statistics on means and standard deviations of the holistic scores per group and draft

		CG		TG		LTG	
		Mean	*SD*	*Mean*	*SD*	*Mean*	*SD*
1	*Adequacy*	1.89	0.33	2.18	0.6	2.4	0.52
	Coherence	2	0.71	2.18	0.41	2.2	0.63
	Cohesion	1.56	0.53	2	0.63	1.9	0.74
	Accuracy	1.67	0.5	1.82	0.6	1.9	0.74
	Mechanics	2	0.71	2.36	0.67	2.6	0.52
	Lexical range	2	0.71	1.91	0.83	2	0.47
3	*Adequacy*	1.78	0.67	2.36	0.67	2.5	0.53
	Coherence	1.78	0.67	2.55	0.69	2.4	0.52
	Cohesion	1.67	0.71	1.73	0.65	1.8	0.79
	Accuracy	1.78	0.67	1.91	0.54	2	0.67
	Mechanics	1.89	0.78	2.45	0.69	2.7	0.48
	Lexical range	1.78	0.67	2.09	0.54	2.2	0.42
4	*Adequacy*	1.67	0.7	2	0.78	2	0.82
	Coherence	2.11	0.6	2.09	0.7	2.5	0.53
	Cohesion	1.33	0.5	1.82	0.6	1.7	0.68
	Accuracy	1.67	0.71	1.73	0.65	2.1	0.57
	Mechanics	2	0.87	2.27	0.65	2.6	0.52
	Lexical range	1.78	0.83	2	0.63	2.1	0.32
6	*Adequacy*	1.78	0.67	1.91	0.7	2	0.47
	Coherence	2	0.7	2	0.45	2.5	0.53
	Cohesion	1.33	0.5	1.36	0.5	2.1	0.57
	Accuracy	1.78	0.67	1.82	0.41	2.2	0.63
	Mechanics	2.22	0.67	2.18	0.87	2.8	0.42
	Lexical range	1.89	0.78	2	0.63	2.5	0.53

The following six tables present the mixed ANOVA results for differences in adequacy, coherence, cohesion, grammatical accuracy, mechanics and lexical range across the drafts written by the pairs in the CG, the TG and the LTG.

Table 63 provides the results for adequacy. The sphericity assumption was met (X^2 = 8.11; p = .151), but apart from the significant main effect for Adequacy ($F_{(3,81)}$ = 3.15; p = .029; η_p^2 = 0.10), there were no statistically significant differences across drafts or between group means, as determined by mixed ANOVA.

Table 63. Mixed-model ANOVA results for adequacy

	Stage 2				
	***df*1**	***df*2**	***F***	***p***	η_p^2
Adequacy	3	81	3.15	.029	0.10
Adequacy * Group	6	81	0.53	.787	0.04
Inter-subjects					
Group	2	27	2.54	.097	0.16

Table 64 displays the output of the mixed ANOVA analysis for coherence. The assumption of sphericity was met (X^2 = 2.98; p = .704), as indicated by Mauchly's test. As can be seen, there is a main effect for Group ($F_{(2,27)}$ = 4.25; p = .009; η_p^2 = 0.18) and for the Coherence x Group ($F_{(6,81)}$ = 3.02; p = .010; η_p^2 = 0.18) interaction.

Table 64. Mixed-model ANOVA results for coherence

	Stage 2				
	***df*1**	***df*2**	***F***	***p***	η_p^2
Coherence	3	81	0.60	.614	0.02
Coherence * Group	6	81	3.02	.010	0.18
Inter-subjects					
Group	2	27	4.25	.009	0.12

Pairwise comparisons showed a statistically significant inter-subject difference. In particular, the TG scored significantly higher than the CG in draft 3 (p = .034, 95% CI [0.05, 1.49], d = 0.50).

For the analysis of cohesion, sphericity was not rejected (X^2 = 1.32; p = .933), and the results, set out in Table 65, reveal a significant main effect for the interaction between Cohesion and Group ($F_{(6,81)}$ = 2.69; p = .020; η_p^2 = 0.17).

Table 65. Mixed-model ANOVA results for cohesion

	Stage 2				
	***df*1**	***df*2**	***F***	***p***	η_p^2
Cohesion	3	81	1.71	.173	0.06
Cohesion * Group	6	81	2.69	.020	0.17
Inter-subjects					
Group	2	27	1.56	.229	0.10

Further post-hoc tests located these differences in the comparison between the first and last drafts written by the TG (p = .009, 95% CI [0.12, 1.15], d = 0.64). Comparisons also found differences between groups, revealing a statistically significant improvement of the LTG's textual cohesion compared both to the TG (p = .010, 95% CI [0.15, 1.32], d = 0.59) and the CG (p = .011, 95% CI [0.15, 1.38], d = 0.58).

Table 66 shows the output with the main results for accuracy. Sphericity was met (X^2 = 4.23; p = .509) and the overall ANOVA did not find a significant difference among means. Thus, accuracy appeared to be unaffected by models when the children's compositions were rated holistically.

Table 66. Mixed-model ANOVA results for accuracy

	Stage 2				
	***df*1**	***df*2**	***F***	***p***	η_p^2
Accuracy	3	81	0.76	.523	0.03
Accuracy * Group	6	81	0.51	.800	0.04
Inter-subjects					
Group	2	27	1.04	.369	0.07

The main results for mechanics are shown in Table 67. Sphericity was not violated (X^2 = 1.56; p = .907). Again, we did not find evidence of a relationship between spelling and models in the global analysis of draft quality.

Table 67. Mixed-model ANOVA results for mechanics

	Stage 2				
	***df*1**	***df*2**	***F***	***p***	η_p^2
Mechanics	3	81	0.28	.843	0.01
Mechanics * Group	6	81	0.77	.599	0.05
Inter-subjects					
Group	2	27	1.75	.362	0.22

The last piece of the output, presented in Table 68, provides the results of the mixed ANOVA for the category of lexical range, whose sphericity was not rejected (X^2 = 8.10;

p = .151). A significant main effect was found for the Lexical range x Group ($F_{(6,81)}$ = 2.57; p = .029; η_p^2 = 0.08) interaction.

Table 68. Mixed-model ANOVA results for lexical range

	Stage 2				
	***df*1**	***df*2**	***F***	***p***	η_p^2
Lexical range	3	81	0.99	.403	0.04
Lexical range * Group	6	81	2.57	.029	0.08
Inter-subjects					
Group	2	27	1.07	.357	0.07

Further pairwise comparisons revealed that the LTG expanded their lexical repertoire significantly in the last writing task as opposed to the first one (p = .049, 95% CI [0, 0.8], d = 1.37).

6.2.1.1 Summary of the main findings for research question 3

The aim of our third research question was to explore whether or not models help improve the written production of primary EFL students in the short and long run. To this end, the children's drafts were first analyzed through quantitative measures of type of clause, complexity, accuracy and fluency and through qualitive analyses following a 3-point scale rubric which evaluated aspects such as adequacy, cohesion, coherence, grammatical accuracy, mechanics or lexical range. The main findings observed in each of the aspects evaluated are the following:

- Type of clause: Within each feedback treatment, the tests indicated a decrease in the production of clauses from the initial to the last version of the texts for the CG. The TG produced fewer pre-clauses on both post-tests. From draft 1 to draft 6, both pre-clauses and proto-clauses decreased significantly. In the case of the LTG, we found significantly fewer pre-clauses from draft 1 to draft 3 and to draft 6, but also fewer proto-clauses from draft 1 to draft 6 as well as more clauses from 4 to 6. Across groups, the treatment groups generated fewer pre-clauses than the CG on the post-test of Cycle 1, and the LTG also gave rise to a

significantly higher number of clauses in draft 6 than the other two feedback conditions.

- Grammatical complexity: After receiving the models in both cycles, that is, on both post-tests, the treatment groups incorporated more subordinate clauses in comparison with their first drafts at Stage 1.
- Lexical diversity: The number of different words used by the learner pairs in the CG and the TG decreased significantly from draft 1 to draft 6. Conversely, the group benefitting from models for four months showed a significant improvement, visible from their first to their last written text. What is more, this group showed higher lexical diversity than their classmates both when they initiated (draft 4) and finished (draft 6) Cycle 2. These results indicate a positive correlation between repeated exposure to models and the variety of words used in the children's texts.
- Accuracy: The only statistically significant result is for the LTG, who made significantly fewer mistakes from draft 1 to draft 6. The difference was also significant with respect to the CG and the TG in drafts 4 and 6, which suggests again that the reception of models for an extended period of time leads to a considerable reduction of errors.
- Fluency: Fluency is the only aspect which underwent a pronounced decline in the three groups. That is, the CG, the TG and the LTG produced significantly shorter texts in their last writing task as compared with the first one.
- Holistic analysis: In the comparison between the first and last texts, we observed that the TG significantly improved the cohesion of their texts, whereas, in line with the quantitative analysis, the LTG enhanced their lexical repertoire. For their part, the children in the self-correction condition did not improve significantly in any of the categories assessed. When the groups were contrasted, the TG seemed to do significantly better than the CG in terms of coherence on the first post-test, while the LTG obtained a significantly higher score in cohesion than the CG and the TG on the second post-test.

In conclusion, do models help improve the written production of primary EFL students in the short and long run? They definitely do. In the short-term, models made the children reduce the number of pre-clauses and proto-clauses and increase the

grammatical complexity of their texts through the use of subordinate clauses. After a long exposure to models, the children were able to (i) produce fewer proto-clauses and more clauses, (ii) use a higher number of different words in their texts and (iii) make fewer errors.

6.2.2 Results for research question 4

The fourth research question is stated here again for the reader's convenience:

4. Does input enhancement play a role in the children's noticing and improvement of third person singular possessives?

In order to explore the extent to which textual enhancement in model texts can have an impact on the knowledge and development of third person singular possessive pronouns among the YLs, the use of this target construction in each group's written text was analyzed. Table 69 shows the summary of the characteristics of the data set.

Table 69. Descriptive statistics on frequency of occurrences, means and standard deviations of third person possessive pronouns per group and draft.

	CG			**TG**			**LTG**		
	FO	*Mean*	*SD*	*FO*	*Mean*	*SD*	*FO*	*Mean*	*SD*
1	18	2	2	40	4.45	1	29	3.2	1
3	23	2.56	2	37	3.82	2	33	3.6	1
4	13	1.44	1	14	1.45	1	14	1.5	1
6	8	0.89	1	13	1.27	1	16	1.8	1

The statistics for the data collected from the CG, the TG and the LTG separately appear in Table 70, Table 71 and Table 72 respectively. As Tables 70 and 71 illustrate, neither the CG nor the TG showed any significant variation across drafts, not even when the first and last compositions were compared. We can see, however, some remarkable tendencies. On the one hand, the percentage of possessive pronouns accurately used by the CG remained quite steady over time, ranging from 62.5% to 76.92%, but slightly decreased in drafts 3 and 6, that is, on the post-tests. Table 70 details these findings.

Table 70. Results of binomial tests for within-group differences in the use of possessive pronouns by the CG

CG	Total	Correct	%	Statistics
Draft 1	18	13	72.22	$z = 0.28$, $p = .776$
Draft 3	23	16	69.56	
Draft 4	13	10	76.92	$z = 0.97$, $p = .332$
Draft 6	8	5	62.5	
Draft 1	18	13	72.22	$z = 0.61$, $p = .539$
Draft 6	8	5	62.5	

Table 71 reveals that the number of correct possessive pronouns written by the TG reached a peak after the treatment, on the first post-test (86.48%). Nevertheless, we can observe a pronounced decline at the beginning of Cycle 2 (64.28%) that remained steady until the end (69.23%). In other words, the children appeared to make more mistakes in Cycle 2 than in Cycle 1.

Table 71. Results of binomial tests for within-group differences in the use of possessive pronouns by the TG

TG	Total	Correct	%	Statistics
Draft 1	40	31	77.5	$z = 1.31$, $p = .190$
Draft 3	37	32	86.48	
Draft 4	14	9	64.28	$z = 0.37$, $p = .709$
Draft 6	13	9	69.23	
Draft 1	40	31	77.5	$z = 0.71$, $p = .475$
Draft 6	13	9	69.23	

In line with the TG, the children in the LTG also showed an improvement in draft 3 (81.81%), as can be seen in Table 72. Nonetheless, this time this rise did not fall over at the beginning of Cycle 2 (92.86%), which suggests that the treatment these pairs received for four months played some role in the development of third person possessives. What is more, the dyads ended up making a correct use of the linguistic target in 100% of the cases, showing a total control of the possessive pronouns. Actually, differences between drafts 1 and 6 were big enough to reach significance.

Table 72. Results of binomial tests for within-group differences in the use of possessive pronouns by the LTG

LTG	Total	Correct	%	Statistics
Draft 1	29	21	72.41	$z = 1.21, p = .226$
Draft 3	33	27	81.81	
Draft 4	14	13	92.86	$z = 1.11, p = .267$
Draft 6	16	16	100	
Draft 1	29	21	72.41	$z = 2.47, p = .013$*
Draft 6	16	16	100	

We also observed some significant differences when the three groups were contrasted. Table 73 displays the statistic results for the differences between groups in terms of proportion of correct answers. As can be seen, only the texts written on the second post-test showed statistical differences. More specifically, it is only the LTG that showed a significant improvement with respect to the other two groups in the last writing task. These results indicate that only with a long exposure to TE were the children able to outperform those who benefitted from a short treatment or no treatment at all.

Table 73. Results of binomial tests for between-group differences in the use of possessive pronouns

	Draft 1		**Draft 3**		**Draft 4**		**Draft 6**	
	z	*p*	*z*	*p*	*z*	*p*	*z*	*p*
CG-TG	0.43	.663	1.59	.111	0.71	.472	0.31	.750
CG-LTG	0.01	.988	1.06	.285	1.16	.244	2.61	.008*
TG-LTG	0.48	.628	0.53	.592	1.84	.065	2.38	.016*

6.2.2.1 Summary of the main findings for research question 4

The objective of the fourth research question was to determine whether the enhancement of possessive pronouns played a role in the children's noticing and use of third person possessives. For this purpose, we examined the learners' texts and analyzed

the children's development in the use of these pronouns. The findings we obtained from this analysis are the following:

- While no statistically significant differences were found across the drafts written by the TG and the CG, we observed some tendencies. Namely, the CG not only did not show any improvement, but the pairs in this condition also became slightly less accurate on the post-tests. The TG provided more correct answers in draft 3, but the dyads' performance turned out to be poorer in Cycle 2. Conversely, the LTG's enhanced noticing of third person possessives in draft 3 remained stable throughout the remaining compositions, reaching a point at which the children used all the third-person possessive pronouns correctly. In addition, statistically significant differences were observed between drafts 1 and 6.
- When the three groups were contrasted, we only observed statistically significant differences between the LTG and the other two groups in draft 6.

Summing up, these findings demonstrate that sustained exposure to highlighted textual linguistic targets had a statistically significant impact on the children's noticing and improvement of third person possessives.

6.3. Module 3: Learners' attitudes and motivation

This third and last module is concerned with the children's motivation as well as their attitudes toward the different feedback techniques. Tapping into young FL learners' attitudes and giving them a voice is key to know whether our task and our assessment is effective and appropriate. In order to provide a more accurate picture than either method would alone, both quantitative and qualitative analyses were conducted. Thus, in the section that follows, we will first present the statistical results obtained from the motivation thermometer. Then, the reasons given in the thermometers as well as the answers provided in the attitude questionnaires and in the interviews will be analyzed qualitatively.

6.3.1 Results for research question 5

The fifth research question is stated here again for the reader's convenience:

5. What is the attitude of students toward the three feedback conditions? Does the students' motivation change throughout the treatment?

In order to look into what level of motivation the learners started from and whether or not their motivation oscillated from pre-task to post-task in each of the two cycles in the three groups, we compared the motivation scores given by each child. Table 74 provides summary statistics for the scores given by each participant in the motivation thermometers.

Table 74. Descriptive statistics on means and standard deviations of the pre- and post-task motivation scores per group and cycle.

	CG (*N* = 18)		**TG (*N* = 22)**		**LTG (*N* = 20)**	
	Mean	*SD*	*Mean*	*SD*	*Mean*	*SD*
			CYCLE 1			
Pre-task	6.94	1.63	7.82	1.30	7.2	1.01
Post-task	7.44	1.58	8.55	1.18	7.9	1.55
			CYCLE 2			
Pre-task	6.61	1.97	7.73	1.24	7.3	1.22
Post-task	8.11	1.41	8	1.54	8.05	1.19

As can be seen, the thermometers indicated that, overall, all learners showed a positive disposition toward the task before doing it (M = 7 for the CG, M = 7.8 for the TG, and M = 7.2 for the LTG). In addition, the post-task scores given by the three groups turned out to be higher in both cycles. At first sight, these rises seem to indicate that the tasks, irrespective of the feedback condition, had the potential to increase the children's initial motivational disposition. Let us see whether the statistical analyses corroborate this assumption.

The results obtained from the mixed ANOVA for the pre- and post-task thermometers are set out in Table 75 below. Closer inspection of the table shows significant effects for Group ($F_{(2,57)}$ = 27.31; p = .001; η_p^2 = 0.24), Cycle 1 ($F_{(1,57)}$ = 6.37; p = .046; η_p^2 = 0.15), and for the interaction between Cycle 1 and Group ($F_{(2,57)}$ = 6.12; p = .049; η_p^2 = 0.08). There also seems to be a main effect for Cycle 2 ($F_{(1,57)}$ = 33.83; p = < .000; η_p^2 = 0.37), and for the Cycle 2 x Group ($F_{(2,57)}$ = 15.4; p = .001; η_p^2 = 0.11), and the Cycle 1 x Cycle 2 x Group ($F_{(2,57)}$ = 3.99; p = .048; η_p^2 = 0.10) interactions.

Table 75. Mixed-model ANOVA results for the within- and across-cycle motivation scores.

	Stage 2				
	***df*1**	***df*2**	***F***	***p***	η_p^2
Cycle 1	1	57	6.37	.046	0.15
Cycle 1 * Group	2	57	6.12	.049	0.08
Cycle 2	1	57	33.83	<.000	0.37
Cycle 2 * Group	2	57	15.4	.001	0.11
Cycle 1 * Cycle 2	1	57	0.66	.420	0.01
Cycle 1 * Cycle 1* Group	2	57	3.99	.048	0.10
Group	2	57	27.31	.001	0.24

Pairwise comparisons revealed that the upward trend mentioned above turned out to be statistically significant for the CG in Cycle 2 (p = .037, 95% CI [-1.45, -1.05], d = 0.58) and for the treatment groups in Cycle 1 (p = .027, 95% CI [-0.09, -1.9], d = 0.58 for TG; p = .003, 95% CI [-0.25, -1.2], d = 0.58 for LTG). The lack of significance in Cycle 2 on the part of the treatment conditions seems to reveal that the comparison to the model texts did not raise their motivation so much. When looking at the learners' motivation

considering pre-task from Cycle 1 against post-task from Cycle 2, only the CG (p = .002, 95% CI [-1.17, -0.28], d = 0.58) and the LTG (p = < .000, 95% CI [-1.47, -0.54], d = 0.58) exhibited significant differences. In other words, only these two groups felt significantly more motivated at the end of the study in comparison with the beginning of it. Therefore, although the three groups manifested a positive shift, the biggest motivation variation was exhibited by the groups that had been working on their corresponding feedback for a longer period of time, namely, the CG and the LTG.

When motivation ratings were contrasted between groups, the pre-task comparisons from both cycles showed that the three groups began Stage 1 with comparable motivation ratings. The motivation scores provided at the post-task of Cycle 2 did not render significant results either. The only significant difference evidenced by the mixed ANOVA tests was found at the post-task of Cycle 1, in which the TG rated their motivation higher than the CG (p = .036, 95% CI [0.17, 1.38], d = 0.58). From these results we could infer that no group felt markedly more or less motivated than the others except for the TG in relation to the CG when they wrote the picture story for the second time.

Regarding motives, the majority of the reasons provided by the children were positive both in the pre-task (25 negative answers vs. 150 positive answers) and the post-task (10 negative answers vs 194 positive answers) thermometers. The reasons selected by the children in the thermometer before and after doing the task in each cycle are summarized in Table 76 and 77, respectively. As the learners could provide more than one reason from the choices available, the percentages presented were calculated based on the total number of children comprising each group. As can be seen in Table 76 below, before the task in Cycle 1, in the CG, 50% expressed that they wanted to work with their partners, 44% anticipated that the task was going to be easy and 28% felt that they were going to have fun doing the task. In the TG, most students (41%) also highlighted the fact that they were going to have fun, but 'I want to work with my peer' (32%) or 'I want to do an activity in English' (32%) were also among the most chosen options. As for the LTG, 45% indicated that they were highly motivated because they wanted to work with their peers, 35% thought they were going to have fun and 30% expressed their desire to do the task. Therefore, at this first stage, differences between

the groups were not noticeable and the children in the three groups chose positive and similar justifications for their scores, these being having fun and working with peers. It is also important to point out that a few children in the treatment groups thought that the task was going to be difficult in contrast with the CG, in which only one learner expressed this fear.

Before doing the task in Cycle 2, the CG once again marked their desire to work with their peers (44%) and thought the task was going to be easy (33%) but this time, instead of choosing the fun component, the children preferred the option 'I want to do an activity in English' (33%). In the case of the TG, it is interesting to observe that not only did no student think that the task was going to be difficult (since they had already done it before), but they also selected the easiness of the task (36%) as the strongest justification for their positive disposition, followed by 'I want to work with my peer' (27%). Moreover, no negative answer was selected this time. In the LTG, 35% expressed that they wanted to work with their peers, 30% wanted to do an activity in English and 30% anticipated that the task was going to be difficult, in contrast with the learners in the TG, who had not selected this option.

In conclusion, it seems that in Cycle 2, the option 'I think I´m going to have fun' decreased considerably in all three groups, which suggests that the children's expectations regarding enjoyment declined as the study progressed. Working with peers keeps on being one of the main reasons for the children's high motivation. The three groups also agreed that they wanted to do an activity in English. However, while the CG and the TG thought the task was going to be easy, for some reason, the children who had been exposed to models for four months thought the opposite. All other motives were marginal with only 1, 2 or 3 learners choosing them.

Therefore, working with peers stands out as the main reason for the scores given at the beginning of the task in both cycles. Let us now see what the children reasoned after doing the task.

Table 76. Mentions and percentages of motivation reasons selected by the three groups at pre-task in Cycles 1 and 2

MOTIVES	**CG (*N* = 18)**		**TG (*N* = 22)**		**LTG (*N* = 20)**	
PRE-TASK **Because I…**	**C1**	**C2**	**C1**	**C2**	**C1**	**C2**
think the task is going to be easy	**8 (44%)**	**6 (33%)**	4 (18%)	**8 (36%)**	3 (15%)	5 (25%)
want to work with my peer	**9 (50%)**	**8 (44%)**	**7 (32%)**	**6 (27%)**	**9 (45%)**	**7 (35%)**
want to do the task	2 (11%)	1 (6%)	6 (27%)	3 (14%)	**6 (30%)**	4 (20%)
want to do an activity in English	1 (6%)	**6 (33%)**	**7 (32%)**	4 (18%)	5 (25%)	**6 (30%)**
think I´m going to have fun doing the task	**5 (28%)**	2 (11%)	**9 (41%)**	3 (14%)	**7 (35%)**	2 (10%)
think the task is going to be difficult	1 (6%)	1 (6%)	4 (18%)	0	5 (25%)	**6 (30%)**
don't want to work with my peer	0	1 (6%)	0	0	0	0
don't want to do the task	0	2 (11%)	0	0	1 (5%)	0
don't want to do the activity in English	1 (6%)	0	0	0	0	0
think I'm going to get bored doing the task	0	1 (6%)	0	0	1 (5%)	1 (5%)

As Table 77 reports, the responses gathered after the task in both cycles were also very positive. Actually, negative answers were barely provided. After the children in the CG had self-corrected their own texts for the first time, having enjoyed working with a peer was the most frequent justification (61%), followed by 'I liked the task' (39%) and 'I enjoyed doing the task' (33%). Thus, it seems that this first task met the children's initial expectations regarding working with their partners and having fun. Like the CG, not only at pre-task, but also at post-task, the TG selected the fact that they had worked with their peers as the most frequent reason for their positive scores (50%). Other most common motives were that the task had been easy (36%), that they had liked the task (36%) and that they had enjoyed doing an activity in English (36%), which also coincides with one of their most common pre-task justifications. As expected, the children in the LTG also highlighted the fact that they had liked working in collaboration (40%). Most of them also considered and that the task had been easy (35%) and some stated that they had enjoyed doing an activity in English (20%) and

that they had enjoyed doing the task (20%). Therefore, like the CG, having worked with their peers and having enjoyed doing the task coincide with their pre-task main motivations.

As in the pre-task results from Cycle 2, the CG still kept a positive attitude toward the task and expressed that they had enjoyed working with their peers (56%) and that the task had been easy (39%). They also added that they had enjoyed doing the task (39%). In the TG, 59% felt that they had enjoyed working with their peers and 23% expressed that they had liked the task. For their part, the learners in the LTG the LTG's fear that the task was going to be difficult in this second stage was resolved, as most of them stated that it had been easy (55%). In addition, they underlined the motives 'I enjoyed working with their peers' (45%), 'I liked the task' (25%) and 'I enjoyed doing the task' (25%).

To sum up, the justifications provided at post-task were mostly positive and common to those selected at pre-task. Overall, these motives reflected that the children found the task easy, enjoyable and motivating due to the possibility to work in collaboration.

Table 77. Mentions and percentages of motivation reasons selected by the three groups at post-task in Cycles 1 and 2

MOTIVES	CG (*N* = 18)		TG (*N* = 22)		LTG (*N* = 20)	
POST-TASK **Because…**	**C1**	**C2**	**C1**	**C2**	**C1**	**C2**
The task was easy	5 (28%)	**7 (39%)**	**8 (36%)**	4 (18%)	**7 (35%)**	**11 (55%)**
I enjoyed working with my peer	**11 (61%)**	**10 (56%)**	**11 (50%)**	**13 (59%)**	**8 (40%)**	**9 (45%)**
I liked the task	**7 (39%)**	3 (17%)	**8 (36%)**	**5 (23%)**	3 (15%)	**5 (25%)**
I enjoyed doing an activity in English	3 (17%)	5 (28%)	**8 (36%)**	4 (18%)	**4 (20%)**	4 (20%)
I enjoyed doing the task	**6 (33%)**	**7 (39%)**	7 (32%)	2 (9%)	**4 (20%)**	**5 (25%)**
The task was difficult	1 (6%)	0	1 (5%)	1 (5%)	2 (10%)	0
I didn't enjoy working with my peer	0	0	0	0	0	0
I didn't like the task	0	0	0	1 (5%)	0	0
I don't like doing activities in English	0	0	0	0	1 (5%)	0
I got bored	0	0	0	0	2 (10%)	0

Apart from selecting one or more motives from the list, some children took the time to write down their own reasons. We considered it important to treat them separately, since they can provide valuable information about motives which were not present among the available options or reinforce some of those already provided. Accordingly, Table 78 shows the justifications added by some learners in the three groups at pre- and post-task in both cycles. The motives presented were each provided by only one student and, although some of them could be merged with those shown above, we preferred to include them all in order to display a faithful and accurate picture.

Table 78. Other reasons provided by the children in the thermometer at pre- and post-task in Cycles 1 and 2

TASK	GROUP	CYCLE	MOTIVE
PRE	CG	1	To learn to work in a team
		1	Because I'm good at English
		1	Because it's fun
		2	Because I know I can do it
		2	Because it's quick to do
	TG	1	Because I like English
		1	Because I get along with my peer
		1	Because I want to learn
		1	Because I have a good time doing tasks
		1	Because I'm going to have a peer to help me
		1	Because if I can't find a word, my peer reminds me of it
		1	Because I'm curious to know what the task is going to be like
POST	LTG	1	Because I have improved my English

This table is quite revealing in several ways. First, all reasons, except for one, were provided at pre-task, and most of them correspond to Cycle 1, which reflects the high expectations and motivation that the children displayed before doing the task for the first time. Another interesting aspect of the data, which supports the results obtained above, is the fact that several students reinforce the value of peer work as a motivating force: 'To learn to work in a team', 'because I get along with my peer', 'because I'm going to have a peer to help me', 'because if I can't find a word, my peer reminds me of it'. From the remaining reasons, two have to do with having fun/having a good time, two deal with being good at/enjoying English and one with curiosity. The only motives shown in Cycle 2 were provided by the CG and were related to the management of the task. The fact that these justifications provided by the CG have not been found elsewhere implies that self-correction was not considered a challenge for them. Actually, this issue seems to be consistent with the motives given by the CG at pre-task in Cycle 2, where 33% of the learners anticipated that the task was going to be easy. Last, but not least, another remarkable aspect to point out is the post-task motive. Although only one learner added this justification, the fact that it comes from the LTG and that the reason this student gives has to do with learning tell us something positive

about the long-term treatment that these children received. It certainly leaves open the possibility that the children from the LTG have actually improved their English.

In conclusion, results concerning motivation reveal an overall positive attitude in the three groups, more positive at post-task than at pre-task, especially for the CG and LTG in which the differences turned out to be statistically significant. No differences were found, however, when the groups were contrasted, except for the relationship between the CG and the TG at post-task of Cycle 1. Regarding motives, the majority of the reasons provided by the children were positive both in the pre-task and post-task thermometers in both cycles. In addition, working in pairs was the main reason chosen to explain their positive disposition, which suggests a role for collaboration in promoting the children's motivation. As a matter of fact, working with their peers was the only motive present in all the motivation questionnaires administered to the three groups and the most common justification added willingly by the children.

Apart from the motivation thermometer, we also gathered some general information about the children's opinion on models and self-correction. This gathering of data comprised two parts: an anonymous questionnaire and focus group interviews. The questionnaire included two sets of questions: the first set contained questions concerning the first stage of the task, that is, writing a text in response to a picture, while the second set was aimed at knowing their opinion about comparing their text with a model/self-correcting their texts.

As for the first set, in response to the question '**do you usually do activities of this type in your English classes at school**?', none of the participants answered affirmatively, which means that the activity of writing a narrative text based on a picture prompt was, at least, not frequently done at their school. Most of the learners in the three groups answered that they had sometimes done this activity, while a minority of participants (33% in the self-correction group and 17% in the TG) answered negatively. Table 79 details these answers.

Table 79. Previous experience with the composition stage in the three groups

	CG (*N* = 6)		TG (*N* = 6)		LTG (*N* = 6)	
	FO	*%*	*FO*	*%*	*FO*	*%*
Yes	0	0	0	0	0	0
Sometimes	4	67	5	83	6	100
No	2	33	1	17	0	0

In line with the level of motivation reported so far, when asked about whether **they had enjoyed the activity** (see Table 80), the majority of the learners in the three groups revealed that they had enjoyed it quite a lot, especially the children in the LTG, who marked the 'quite a lot' option in 83% of the cases. In addition, these children did not report any negative answer in comparison with the other two groups.

Table 80. Level of enjoyment of the composition stage in the three groups

	CG (*N* = 6)		TG (*N* = 6)		LTG (*N* = 6)	
	FO	*%*	*FO*	*%*	*FO*	*%*
Not at all	1	17	0	0	0	0
Not so much	0	0	1	17	0	0
So-so	2	33	1	17	1	17
Quite a lot	3	50	3	50	5	83
Very much	0	0	1	17	0	0

Among the negative responses gathered, the participants' main reasoning was that the task was boring and tiring (see examples 11-14). Furthermore, one participant confessed to feeling uncomfortable being recorded (see example 15). On the contrary, those who valued the activity positively explained that they had liked the task (16-17), had felt comfortable doing it (18), had missed class (13, 19 and 26) or had fun (20 and 27). However, in accordance with the motives reported in the motivation thermometer, many learners referred to working with their peers as the main reason for having enjoyed the activity (21-23). It is also interesting to observe that just over half of the children in the LTG based their enjoyment on learning (24-27).

(11) *we have done it too many times and it is very tiring* (CG)

(12) *it was a bit boring* (CG)

(13) *it was boring, but you missed class* (TG)

(14) *it's not what I like the most in the world, but it was good* (TG)

(15) *we were recorded* (CG)

(16) *I liked it* (CG)

(17) *it has not been that bad* (TG)

(18) *I have felt comfortable doing the activity* (TG)

(19) *I have missed class* (TG)

(20) *It was fun* (LTG)

(21) *I worked comfortably with my peer* (CG)

(22) *I find it fun to do activities with my classmates* (CG)

(23) *I have worked well with my partner* (LTG)

(24) *I think it's good that we do things in English* (LTG)

(25) *I have done an activity in English* (LTG)

(26) *We have learned (and missed class)* (LTG)

(27) *It was fun, and we improved our English* (LTG)

When students were asked their opinion about **the usefulness of the activity to learn English**, we can see in Table 81 that the overall response to the question was very positive. The majority of the children in the CG (67%) considered it to be quite useful, while the responses of the treatment groups were divided between ‘quite a lot’ and ‘very much’. It is important to underscore that 66% of these positive answers belong to the TG in opposition to an 83% which corresponds to the LTG. Therefore, once again, the

children who had benefitted from the treatment, especially the children from the LTG, seemed to be more positive than those who had not received any feedback.

Table 81. Level of usefulness of the composition stage in the three groups

	CG (*N* = 6)		**TG (*N* = 6)**		**LTG (*N* = 6)**	
	FO	*%*	*FO*	*%*	*FO*	*%*
Not at all	0	0	0	0	0	0
Not so much	1	17	1	17	0	0
So-so	0	0	1	17	1	17
Quite a lot	4	67	2	33	3	50
Very much	1	17	2	33	2	33

As for the reasons provided, a note of caution is due here, since the children tended to give their opinion on the whole task rather than on the first stage. In consequence, what we will encounter here are their general feelings about the entire treatment. For example, one student from the CG felt that self-correction was not enough (28), and another learner reasoned that the models were too easy for them (29). Apropos of the positive responses, the children found the activity useful mainly because it improved their written and spoken English (30-33) or because they learned some other aspects (34-39). Surprisingly, one student also commented on having learned to use the third person possessive pronouns (40).

(28) *We didn't work on the texts after self-correcting them* (CG)

(29) *I already knew everything* (TG)

(30) *We write in English* (TG)

(31) *We speak in English* (CG)

(32) *We practice speaking* (LTG)

(33) *You learn to write in English* (TG)

(34) *You learn more things than just writing and speaking* (TG)

(35) *Thanks to this we are able to do more things in English* (CG)

(36) *It helps you know how to express what you see* (LTG)

(37) *You learn a lot of things* (LTG)

(38) *I have learned new things* (LTG)

(39) *It improves our learning* (TG)

(40) *I have learned to use his/her* (LTG)

To finish with the first set, the students were asked about **whether they would like their teacher to continue doing this type of activity**. Although it has been apparent from the beginning of the questionnaire that the CG was not as favorable to self-correction as the treatment groups were to models, this may be the first question in which we can really tap into the children's real feelings toward the task. As can be seen in Table 82, 67% of the participants in the CG did not want their teacher to implement this activity, while over half of the children in the TG and the totality of the learners in the LTG answered positively.

Table 82. Participants' eagerness to write a text based on a picture prompt in future lessons

	CG (*N* = 6)		**TG (*N* = 6)**		**LTG (*N* = 6)**	
	FO	*%*	*FO*	*%*	*FO*	*%*
Yes	2	33	4	67	6	100
No	4	67	2	33	0	0

While the CG's negative justifications show indifference and boredom (41-44), the TG's main fear was making mistakes (45) and feeling uncomfortable with a classmate (46). The positive answers include learning (47-50), missing class (51), having fun (52-53) and other motives (54-56). As mentioned above, these answers start to reflect an unenthusiastic view toward self-correction, and a positive attitude toward models, especially for those who were exposed to this feedback technique for a longer period of time. Let us see what the children answered when asked directly about self-correction and model texts.

(41) *It's very tiring* (CG)

(42) *We have enough with the English classes* (CG)

(43) *It has been boring* (CG)

(44) *I don't know, I didn't like it very much* (CG)

(45) *I may make mistakes* (TG)

(46) *I may work with a partner I don't get along with* (TG)

(47) *That's how we learn* (CG)

(48) *I want to learn* (TG)

(49) *We would learn more* (LTG)

(50) *That's how we improve our level of English* (LTG)

(51) *You miss class* (TG)

(52) *I would have a good time* (TG)

(53) *It's fun* (LTG)

(54) *From time to time it's okay* (TG)

(55) *It's interesting and entertaining* (LTG)

(56) *It's useful for us* (LTG)

The second section of the questionnaire required respondents to give their opinion on the comparison or self-correction stage by answering the same questions posed above. Table 83 and Graph 5 feature the participants' answers to the question '**do you usually do activities of this type in your English classes at school?**'. On the one hand, the table below illustrates that opinions differed as to whether or not self-correction was implemented in their classes, as half of the focus group from the CG answered 'sometimes' and half of them answered 'no'. We could therefore surmise that they were familiar with the task and that, at least at some point, they self-corrected their own texts

in their English classes. On the other hand, almost all children in the treatment groups agreed that models had sometimes been used as a feedback technique by their teachers.

Table 83. Previous experience with models/self-correction in the three groups

	CG (*N* = 6)		**TG (*N* = 6)**		**LTG (*N* = 6)**	
	FO	*%*	*FO*	*%*	*FO*	*%*
Yes	0	0	1	17	0	0
Sometimes	3	50	4	67	6	100
No	3	50	1	17	0	0

Graph 5. Previous experience with models/self-correction in the three groups

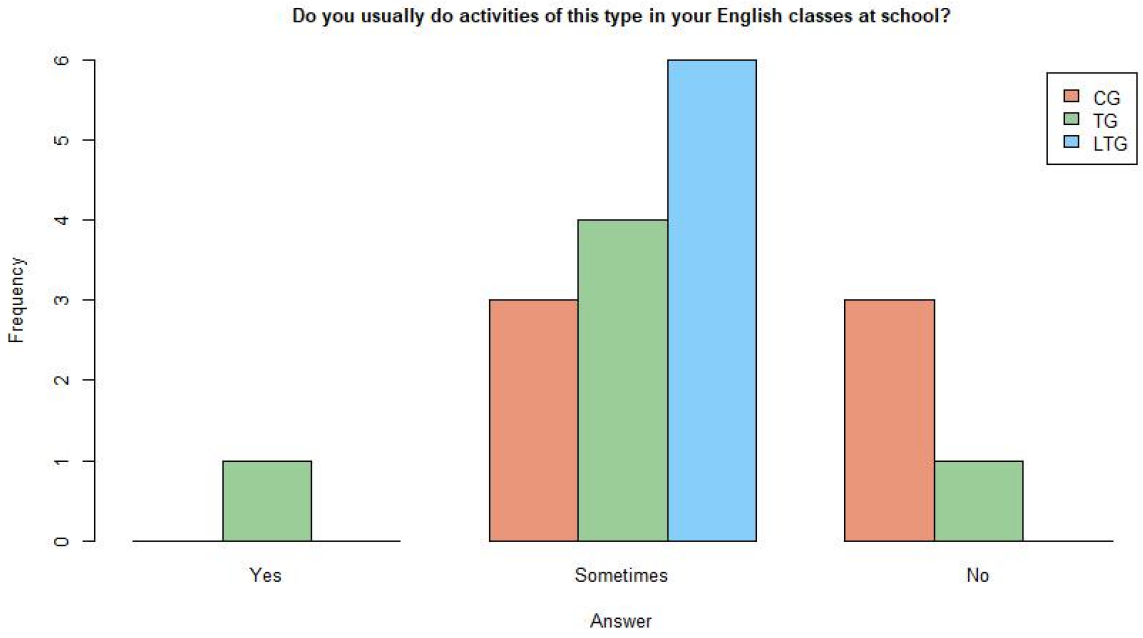


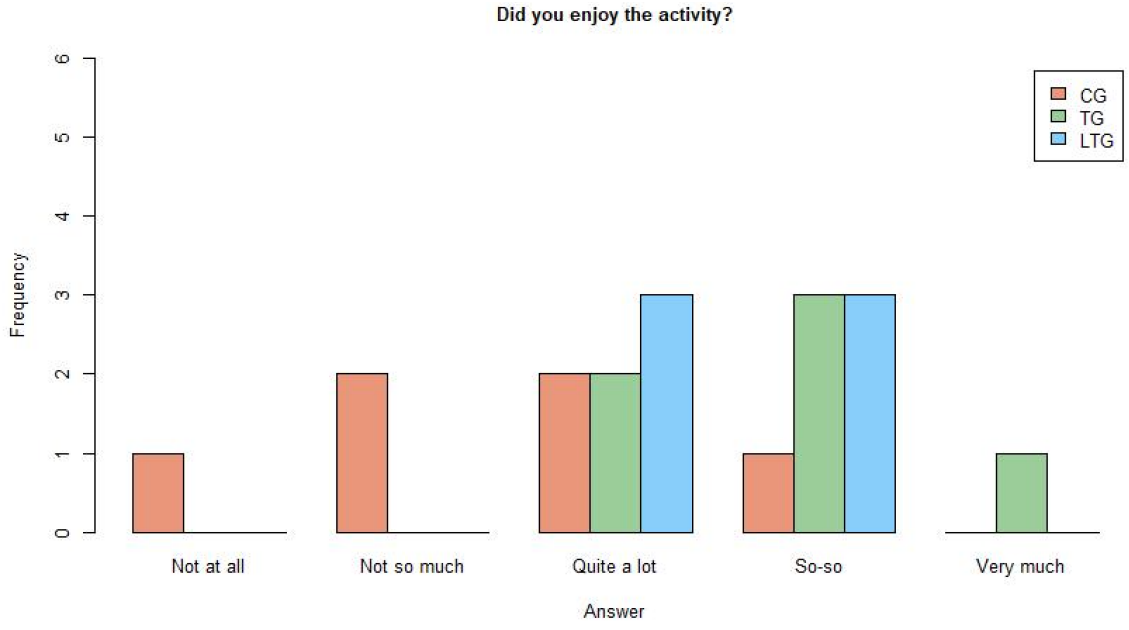
Table 84 and Graph 6 present the children's responses when canvassed about their **enjoyment doing their corresponding activity**. The CG's opinions differed. That is, half of the respondents from the CG considered that they had not enjoyed it as much or at all and half of them marked the 'so-so' and 'quite a lot' options. Models, by contrast, spurred motivation for the task and elicited positive responses. Specifically, most of the TG's responses ranged from 'so-so' to 'quite a lot' and one child from the TG indicated their full enjoyment of the comparison activity. In the case of the LTG, the majority of those who responded to this item expressed that they had enjoyed the task quite a lot,

which corroborates the answers provided in the previous sections, thus pointing to a gradual liking for the task: The monotony of self-correction led to an overall pessimistic feeling toward the task, while the use of models in both cycles elicited positive responses from the learners, and even more so in the case of the children who benefited from the models over a longer period of time.

Table 84. Level of enjoyment of the comparison or self-correction stage in the three groups

	CG (*N* = 6)		**TG (*N* = 6)**		**LTG (*N* = 6)**	
	FO	*%*	*FO*	*%*	*FO*	*%*
Not at all	1	17	0	0	0	0
Not so much	2	33	0	0	0	0
So-so	1	17	3	50	1	17
Quite a lot	2	33	2	33	5	83
Very much	0	0	1	17	0	0

Graph 6. of in the three groups



Again, a common view amongst the children in the CG was that they found the activity boring (57-58). Some other informants also alluded to the notion of being recorded (answer probably provided by the same participant who gave this answer three

questions above) and one learner also suggested that they were used to being corrected by the teacher (60). As for the positive feedback, a variety of perspectives were expressed: working collaboratively (61-62), having fun (63-64), missing class (65), feeling comfortable (66), enjoyment (68) and learning (69-71). All this data suggests that the children's attitudes toward a given form of feedback are affected by many factors other than the nature of the technique itself.

(57) *It's very, very, very boring* (CG)

(58) *It bored us* (CG)

(59) *We have been recorded* (CG)

(60) *I'm used to being corrected by the teacher* (CG)

(61) *I have worked comfortably with my partner* (CG)

(62) *I find it fun to do activities with my colleagues* (CG)

(63) *I had a good time* (TG)

(64) *It was fun* (LTG)

(65) *I missed class* (TG)

(66) *I have felt comfortable* (TG)

(67) *I like English* (LTG)

(68) *I liked it* (LTG)

(69) *I learn vocabulary* (LTG)

(70) *It's useful to compare* (LTG)

(71) *We have learned from our mistakes in writing* (LTG)

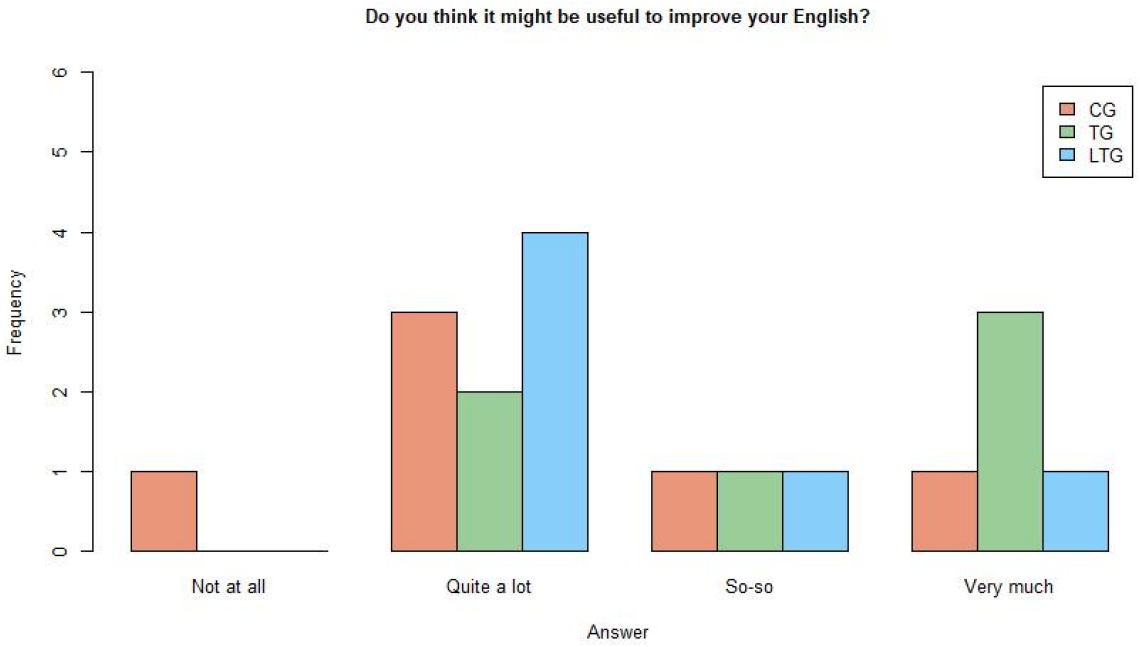
On the topic of the usefulness of modelling and self-correction, Table 85 below reveals that, irrespective of the feedback condition they were assigned to, most learners found

both techniques helpful to improve their English, which reflects that, although many of the informants, especially from the CG, reported not having enjoyed these activities, they acknowledged their usefulness to learn the language.

Table 85. Level of usefulness of the comparison/self-correction activity in the three groups

	CG (*N* = 6)		**TG (*N* = 6)**		**LTG (*N* = 6)**	
	FO	*%*	*FO*	*%*	*FO*	*%*
Not at all	1	17	0	0	0	0
Not so much	0	0	0	0	0	0
So-so	1	17	1	17	1	17
Quite a lot	3	50	2	33	4	67
Very much	1	17	3	50	1	17

Graph 7. Level of usefulness of the comparison/self-correction activity in the three groups



The only participant who did not consider the activity to be helpful thought that self-correction is an activity in which they can easily forget what they learned (72).

However, the positive justifications were abundant and are certainly in line with the afore-mentioned reasons (73-80).

(72) *We forget everything after a while* (CG)

(73) *We can learn* (CG)

(74) *It helps us know what we are doing wrong* (CG)

(75) *We speak in English* (CG)

(76) *You learn to write in English* (TG)

(77) *You learn more things* (TG)

(78) *You learn more and have fun* (LTG)

(79) *To know how words are spelled* (LTG)

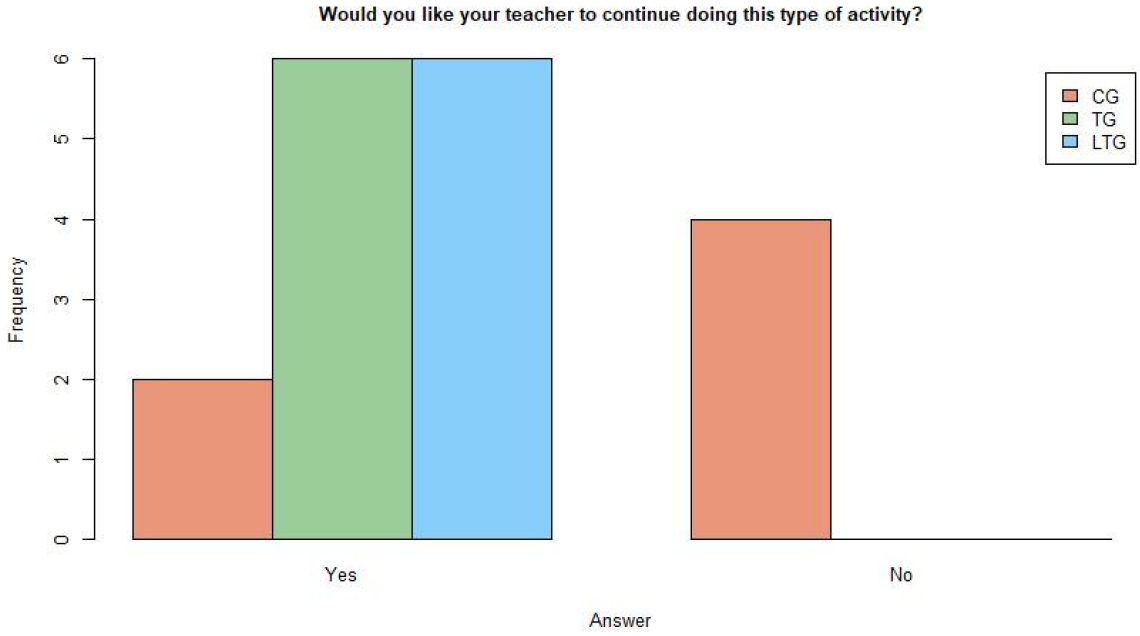
(80) *I have learned new things* (LTG)

Moving on to the last and most revealing question, when the students were asked **whether they would like their teacher to continue doing the corresponding activities**, they answered very differently, as can be seen in Table 86 and in Graph 8. Whereas all the participants in the treatment groups were unanimous in the view that models should be implemented in the classroom, 67% of the learners in the self-correction group thought otherwise. These answers certainly tell us about the significance of considering the motivation factor when thinking about implementing new tasks or approaches in the EFL classroom. As helpful as these activities may be, if children are not motivated, the results may not be as expected.

Table 86. Participants' eagerness to use models/self-correcting in future lessons

	CG (*N* = 6)		**TG (*N* = 6)**		**LTG (*N* = 6)**	
	FO	*%*	*FO*	*%*	*FO*	*%*
Yes	2	33	6	100	6	100
No	4	67	0	0	0	0

Graph 8. Participants' eagerness to use models/self-correcting in future lessons



All the reasons given in favor of the continuity of these techniques had to do with learning (81-89) and enjoyment (90-91). Those against the continued use of self-correction claimed not to like the task (92-94) and prefer to receive feedback (95).

(81) *It helps us with our English* (CG)

(82) *That's how we learn* (CG)

(83) *It is useful* (TG)

(84) *Sometimes it comes in handy* (TG)

(85) *I want to learn* (TG)

(86) *If there is someone who is bad at English, I think this would help* (LTG)

(87) *To learn more* (LTG)

(88) *To practice* (LTG)

(89) *We improve our level of English* (LTG)

(90) *I liked it* (LTG)

(91) *I enjoyed the activity* (LTG)

(92) *I don't like this activity at all* (CG)

(93) *I don't like these kinds of activities* (CG)

(94) *I didn't like it very much* (CG)

(95) *I prefer to be corrected by the teacher* (CG)

The last section of the questionnaire required our 18 respondents to add or comment freely about any aspect they wanted to mention. This is probably the most interesting question, as it gave the children total freedom to express themselves without being constrained by specific questions, allowing us to get a real sense of what our young participants really thought. For example, the informant who provided the first answer below (96) insinuated boredom, tiredness, and exasperation at doing these tasks when they could devote their time to more useful activities. This view was partially echoed by another informant (97) who considered self-correction ineffective in comparison with the teacher's correction, to which they are used to. By contrast, the learner in 98 did not completely stigmatize self-correction given its usefulness to learn how to correct their own mistakes, but they would not do this activity over. In relation to the TG, two participants (99-100) revealed their desire to do the activities more frequently and another child expressed being satisfied with the task (101). Three learners from the LTG also left their thoughts in this comment section and put forward three different aspects. One learner wanted to highlight the fact that he had liked the task (102), while another student (103) suggested that doing the task in one or two stages rather than in four stages would be more appropriate. In line with the TG, the third child (104) manifested their desire to do the activities more often.

(96) *This activity is a real drag. It annoys me a lot when we are interrupted in class while we are doing activities. Besides, we always do the same and I'm so tired!* (CG)

(97) *I don't find it very educational to correct our things, when it's normally the teacher who corrects us* (CG)

(98) *I wouldn't like to repeat it again, but, if necessary, it is not bad to improve our English, or to learn more and know how to correct our mistakes* (CG)

(99) *I liked the task a lot and it should be done more frequently* (TG)

(100) *We should do this activity more often* (TG)

(101) *Well, I think it's fine the way it is* (TG)

(102) *I liked it a lot* (LTG)

(103) *I would like to do the whole process in one or two parts instead of doing the process in different parts* (LTG)

(104) *We should do this more often* (LTG)

Overall, these answers capture the gist of what we have seen throughout the questionnaire. More especifically, some of the children in the self-correcton condition recognized that this technique might be useful to improve their English and learn how to correct their mistakes, while some others considered self-correction a worthless activity and gave more value to external feedback. On the other hand, all of the participants in this group seemed to agree with the feeling that correcting their own texts was monotonous, boring and tiring, which fully contradicts the scores provided in the motivation questionnaire. On the contrary, the treatment groups, especially the LTG, broadly valued the use of models positively, not only in terms of learning, but also with regard to enjoyment. Figure 4 below provides a visual representation of the most frequent words employed by the children in the treatment groups throughout the questionnaire. The larger the term, the higher the frequency. Therefore, as can be seen, 'learn', 'fun' and 'English' are the most frequent words found in our database. Closer inspection of the word cloud shows that most of the terms used are positive ('liked', 'interesting', 'practice', 'good', 'comfortable', 'enjoy', 'well', 'improve') and others exhibit the practical side of the feedback technique ('mistakes', 'vocabulary', 'speaking', 'writing', 'different', 'handy', 'partner', 'spelled', etc.).

Figure 4. Graphical representation of word frequency in the treatment groups

In the final part of the motivation and attitudes survey, the three focus groups were interviewed with the purpose of obtaining a collective response on several issues, which allowed us to attain deeper knowledge of some of the aspects raised above and learn about new ones. One of the questions worth exploring is the issue of learning. That is, when the children were asked what they thought the respective task was useful for, the richness of the answers varied across groups. Among the responses provided by the CG we find rather general answers: '*It's one more activity to practice English*', '*to have more knowledge of English*', '*to learn how to write texts*'. The participants in the TG also gave vague answers such as '*to learn more*' or '*to improve our English*', but one interviewee added a more specific reasoning: '*to improve our English and learn vocabulary*', an answer with which their classmates agreed. The children in the LTG, on the other hand, elaborated on their answers in more depth:

(105) Excerpt of conversation:

> CHI48: *We have learned from our mistakes, and ... you learn more things than you knew*
>
> CHI57: *Also the words ... how they are pronounced, how they are written ...*

CHI59: *'His' and 'her' too*

CHI55: *And knowing how to express yourself, right?*

CHI57: *Yes*

CHI48: *Yes*

CHI55: *Learning how to write what you see*

CHI57: *And also talking to our peers and helping each other. In general, I think I have improved my English*

CHI55: *Yes*

CHI41: *Yes*

CHI48: *I think so*

CHI59: *Yes*

CHI51: *Yes*

CHI48: *But overall, my vocabulary*

CHI51: *Yes, vocabulary*

CHI55: *Vocabulary*

CHI57: *And also 'his' and 'her' ... before ... it was worse for me*

CHI55: *And learn the words*

CHI41: *And make sentences*

CHI48: *Yes*

As we can see, the responses given by the LTG were wide-ranging. The children were able to verbalize a broad number of issues that they had learned according to their own ideas and even a couple of them raised the issue of the third person possessive pronouns, which is surprising given the implicit nature of this technique.

The next question was aimed at eliciting information about how they would like to be corrected. All the participants in the CG did not like correcting their own texts and expressed their preference for the teacher marking the mistake without providing the correct form. In response to the question 'would you like to receive a native model as a

feedback technique?', they all answered affirmatively. The TG, on the other hand, argued that they would have liked to get the models immediately after writing the texts:

(106) Excerpt of conversation:

CHI32: *We would have liked to be corrected as soon as we finished*

CHI22: *Yes, that you don't have to wait for the next session to receive the model*

Despite the answers above and in line with the CG, when they were asked to choose between receiving models or being corrected explicitly, they opted for the latter:

(107) Excerpt of conversation:

CHI19: *This way you learn more, and you are not ...*

CHI22: *You can immediately see the mistake on the paper, and you don't need to be comparing ... which gets yourself into a mess*

When the LTG was asked about this issue, their answers were very much consistent with those provided by their classmates:

(108) Excerpt of conversation:

CHI41: *Actually, I prefer the teacher to circle the mistake and let us think*

CHI57: *Yes*

CHI48: *That's it*

CHI41: *Otherwise, we don't learn ... If they tell us everything ...*

CHI57: *Yes, that's it. They mark it for us and then we spend some time correcting everything*

CHI48: *But the models were pretty good. And they were easier because you can remember some things for the next text*

CHI55: *It's easier, because if you don't know how to say something, it will most likely appear in the text*

To summarize, a common view amongst interviewees in relation to this question was that they favored a more explicit form of error correction (through the indication that an error has been committed) over self-correction and models. Nevertheless, a number of other issues were identified. For instance, the children in the CG showed a preference for receiving models over self-correcting their texts, while those in the TG seemed to prefer more immediacy in the reception of models. Finally, the LTG regarded models as a feedback technique which enabled them to retain information more easily.

The following question asked them to specify what they liked and disliked about the activities. For the CG, working with a peer was the most appealing aspect of the task (although some of them complained about their partners). On the other hand, negative concerns were expressed about the task being too repetitive, while some students suggested exchanging their texts and correcting one another's texts instead of their own. In the same way, the children in the TG liked working with their partners, but also reported having fun with the task:

(109) Excerpt of conversation:

CHI19: *Well, I liked it because I have been working very well with my partner and I have had a good time*

CHI35: *I liked it because writing the text seemed funny to me because of... I don't know... because afterwards we had to compare it with a text*

CHI32: *Yes, yes, I liked it because I worked with my friend*

Concerning their dislikes, there seemed to be a sense of shyness among the children, as they unanimously expressed feeling uncomfortable being recorded. One respondent also disapproved of having to repeat their writings:

(110) Excerpt of conversation:

CHI22: *Well ... I don't know*

CHI19: *Well, let's see ... Well, what ... Well, I ... I liked everything*

CHI22: *Me too. I don't know ... I haven't been upset, really*

CHI32: *I didn't like being recorded*

CHI39: *I didn't like repeating what I first wrote*

CHI22: *The first time you recorded us I felt very shy*

CHI32: *We are much shyer and that is why we speak low, because we don't want to ... I mean, it's not that we don't want to, it's that we're shy*

CHI39: *Yes*

CHI25: *That's it*

CHI22: *It's like we're in our privacy, so the first time it kind of intimidates you*

The participants in the LTG also agreed that working in collaboration was the best part along with missing class:

(111) Excerpt of conversation:

CH48: *Missing class*

CHI57: *And ... well… yes, working with our partner and helping each other*

CHI48: *That's it*

CHI55: *Yes, working with a partner*

Although there were no negative comments, the learners suggested starting with the task from the beginning of the course on the grounds that (i) they feel more tired in the second term and (ii) they would feel more prepared for the next educational stage:

(112) Excerpt of conversation:

CHI48: *No*

CHI57: *Anyway ... I would have started at the beginning of the year*

CHI48: *Yes*

CHI55: *But that's what we did*

CHI57: *But not since the beginning*

CHI48: *But we started in the second term*

CHI41: *But now we are more tired than at the beginning of the year*

INTERVIEWER: *You mean, at the beginning of the course?*

CHI57: *Yeah, well, we are starting to improve now and it would be nice to start at the beginning of the course because we wouldn't be so tired and we are also moving up to secondary school and we would do better*

CHI48: *Yes*

All the children without exception showed their content at having performed the task with a classmate, which is consistent with the reasons given in the motivation thermometer. The justifications for not liking the task are not new either: both the CG and the TG complained about the task being too repetitive, and once again the TG brought up the recording issue as a negative side of the task. Additionally, two suggestions were offered by the CG and the LTG: the CG proposed correcting other

pairs' written texts to make the task more amenable, and the LTG leant toward initiating the experiment with models at the beginning of the school year.

We were also interested in knowing in which aspects the children considered they had improved. As can be seen in the excerpt below, in almost all cases, the informants in the self-correction group reported having enhanced their spelling. A couple of respondents indicated that they had learned a bit of everything or had increased their lexicon as well as improved their speaking skills:

(113) Excerpt of conversation:

CHI18: *In everything*

CHI8: *In spelling*

CHI11: *In spelling*

CHI13: *In speaking*

CHI11: *In spelling, because when you talk to your partner to see how you write a word, you learn, because he knows one thing and you know another.*

CHI8: *Spelling*

CHI15: *Vocabulary*

CHI1: *Spelling and vocabulary*

CHI18: *I … in a bit of everything*

In the case of the treatment groups, the answers provided were more widespread. Most of them pointed to their improvement in vocabulary, but some other interviewees added that they had enhanced their grammar, writing skills and improved the use of the third person possessive pronouns.

(114) Excerpt of conversation (TG):

CHI19: *In vocabulary*

CHI32: *In vocabulary*

CHI25: *Yes*

CHI39: *Yes, in vocabulary*

CHI19: *And in grammar*

CHI25: *And writing*

CHI35: *And in the use of 'his' and 'her'*

CHI22: *And we also learned how to write a text in English*

(115) Excerpt of conversation (LTG):

CHI41: *Yes, in vocabulary*

CHI48: *Yes, in vocabulary, grammar and spelling*

CHI57: *Yes, and in the use of 'his' and 'her'*

The last question aimed to find out if the children had answered the first motivation questionnaire honestly or if, on the contrary, their motivation faded away as the experiment progressed. Therefore, to the question 'have you kept the same level of motivation throughout the entire process?', opinions differed depending on the feedback condition. The participants in the CG definitely lost their motivation at some point, as the intervention below illustrates:

(116) Excerpt of conversation:

CHI15: *We have lost it*

CHI8: *We have lost it, yeah*

CHI18: *The whole process*

CHI18: *I do*

CHI15: *I don't*

CHI8: *It has flown away*

CHI11: *I liked it*

CHI13: *No way!*

The learners in the treatment groups provided more positive answers, although some of them admitted occasionally feeling less eager to do the activity:

(117) Excerpt of conversation (TG):

CHI19: *I haven't lost it, but the first day I was a bit confused, because they had recorded me and I thought… I don't know, I mean… But then I got used to it and I liked it…*

CHI35: *I didn't like it at all the first day, but then I started to like it*

CHI22: *Yes*

CHI32: *I haven't lost it, but sometimes I didn't feel like writing*

(118) Excerpt of conversation (LTG):

CHI48: *Weeeell*

CHI51: *I have been very motivated from the beginning*

CHI41: *Well…*

CHI55: *I don't know*

CHI57: *At the beginning it was fine, but some days...*

CHI41: *There were days when you thought 'I'm feeling way too lazy to write!'*

CHI48: *Yes*

CHI57: *Yes. But ... sometimes you felt like doing something different and working with your classmate*

CHI48: *Depending on the day, you can feel very lazy*

CHI57: *Especially on Mondays*

6.3.1.1 Summary of the main findings for research question 5

Research question 5, which in turn closes the third and last module of this results section, aimed to explore the children's attitudes toward the three feedback procedures as well as measure their motivation before and after performing the task in both cycles. To this end, three questionnaires were administered to the participants. The main findings obtained from the overall answers provided were the following:

- Results concerning the children's motivation revealed an overall positive attitude in the three groups, more positive at post-task than at pre-task in the two cycles, especially for the CG and the LTG in which the differences turned out to be statistically significant when Cycle 1 and 2 were compared. However, this upward shift did not turn out to be so apparent, as subsequent answers did not match those provided in the motivation thermometer, thus uncovering the real children's disposition. More precisely, the self-correction group showed a downward fluctuation in motivation, while the treatment groups manifested a more positive attitude, especially the LTG. When the groups were contrasted, no differences were found except for the relationship between the CG and the TG at post-task of Cycle 1.
- Regarding motives, most of the reasons provided by the learners were positive both in the pre-task and post-task thermometers. Furthermore, social justifications were the most frequently selected motives to explain their positive

motivational disposition, which suggests a role for collaboration in promoting the children's motivation. As opposed to the scores, this justification surfaced in the three questionnaires.

- On the question of the participants' previous experience with self-correction or models, the CG reported being familiar with the task, and almost all children in the treatment groups agreed that models had sometimes been used as a feedback technique by their schoolteachers.
- When the students were asked whether they would like their teacher to continue doing the corresponding activities, all the participants in the treatment groups agreed that models should be implemented in the classroom, while most of the learners in the self-correction group thought differently.
- In general terms, some of the children in the self-correction condition acknowledged that this form of feedback might be useful to improve their language skills and learn how to correct their mistakes, whilst some others regarded self-correction as a worthless activity and placed more value on external feedback. Nevertheless, all of the participants in this group seemed to share the view that correcting their own texts was monotonous, boring and tiring, to say the least, which contravenes the scores obtained in the motivation thermometer. In opposition, the treatment groups, especially the LTG, by and large valued the use of models positively, not only in terms of learning, but also with regard to enjoyment. All this data implies that young learners' attitudes toward a given form of feedback are affected by many factors other than the nature of the technique itself.

Consequently, we could conclude in broad terms that the monotonous and unrewarding nature of self-correction resulted in an overall pessimistic feeling toward the task, although the children admitted its usefulness for learning spelling and vocabulary, among other aspects. By contrast, the use of models in both cycles elicited positive responses from the learners, and even more so in the case of the children who benefited from the models over a longer period of time. Although many participants expressed their preference for a more explicit type of error correction (marking the mistake without providing the correct form), their enjoyment, improvement in vocabulary, grammar, writing skills and in the use of third person possessives alongside their high

motivation to work with their peers certainly make models a powerful pedagogical tool to implement occasionally in the EFL classroom.

CHAPTER 7: DISCUSSION OF MAIN FINDINGS

The main goal of the present dissertation was to investigate the lasting benefits of modelling on children's L2 development in depth. Specifically, this dissertation intended to explore the impact that child oral interaction has on their written product; obtain evidence on whether long-term engagement with writing practice and feedback can bring about learning; analyze the effect of input enhancement on the students' use of third person singular possessive pronouns as well as measure the learners' motivation along the way. In the following sections, a summary of the main findings of this study will be provided and interpretations will be made in relation to relevant theory and research. For ease of reading, each research question will be restated and discussed in conjunction with the hypotheses entertained in Chapter 5 (pages 131-137).

This chapter will consist of three sections that correspond to the three modules in which both the research questions and results have been organized. In section 7.1, the results concerning the oral-written connection obtained from the analysis of the students' engagement with language and content-related episodes and its impact on their written performance will be summarized and discussed in relation to Hypotheses 1 and 2. Section 7.2 will deal with the findings from the examination of the learners' written performance to explore Hypotheses 3 and 4. Finally, the study of the children's motivation will be addressed in Section 7.3 to examine Hypothesis 5.

7.1. Oral-written connection

Research Question 1

What linguistic features do young EFL children focus on when dealing with a task using model texts?

a) *What features of language do they notice when composing a narrative text?*
b) *What do they notice when comparing their written texts with model texts or when correcting themselves?*

c) *Are there any across-cycle differences within the three groups (CG, TG and long-term treatment group (LTG)) regarding frequency, type and outcome of the episodes produced in oral interaction?*

Research question *1a* sought to determine which features child EFL noticed the most or found most problematic when writing a text in collaboration in cycles 1 and 2. As for Cycle 1, no between-group differences were detected in the total number of features noticed while writing a text in collaboration, and the three groups attended mostly to lexis (verbs and nouns), form (verb form and verb tense) and mechanics (spelling and punctuation) rather than to discourse or CREs. When the three groups were contrasted, no statistically significant differences were found for any of the categories analyzed. This suggests that the participants assigned to either of the three groups were similar with respect to the amount and nature of the PFNs. Hypothesis 1a was supported as seen in Luquin and García Mayo (2021). Our results, however, are not in line with those obtained in research conducted with both EFL child (Cánovas Guirao et al., 2015; Coyle & Roca de Larios, 2014, 2020; Coyle et al., 2018; Lázaro-Ibarrola, 2021; Luquin & García Mayo, 2020, 2021; Villarreal & Lázaro-Ibarrola (submitted)) and adult learners (Abe, 2008; García Mayo & Loidi Labandibar, 2017; Hanaoka, 2006a, 2006b, 2007; Hanaoka & Izumi, 2012; Kang, 2020; Martínez Esteban & Roca de Larios, 2010; Montealegre Ramón, 2019; Yang & Zhang, 2010), which revealed that at this first stage of the writing process noticing was mainly lexically driven.

Our young participants' main concern had to do with grammatical issues, in contrast with findings in previous studies which showed that grammar received less than 5% of attention (e.g., Cánovas Guirao et al., 2015; Coyle & Roca de Larios, 2014). Nevertheless, even with the addition of a third group, our results are aligned with Luquin and García Mayo (2020, 2021), who also observed that formal aspects were in the lead of the PFNs at Stage 1. There are two likely causes for this overriding attention to grammar. First, as indicated by Luquin and García Mayo (2020, 2021), this finding might be ascribed to how grammar is still presented to YLs in the FL classroom. Even now, many teachers work almost exclusively with a grammar-oriented approach or rather a FonFs practice. Throughout their school lives, children sit for many official exams that evaluate their linguistic accuracy over other more communicative language skills. Therefore, we may speculate that our participants faced this first task as one more

type of assessment and, as a result, formal aspects of the language such as grammatical accuracy, spelling or punctuation came to the fore, leaving lexis in a second place. As proposed by Cánovas Guirao (2017) and Coyle and Roca de Larios (2018), it also seems possible that this primary focus on formal aspects is attributed to the training session which may have acted as a catalyst for the allocation of the children's attentional resources to other linguistic aspects different from lexis, at least at this first stage. Our young participants were guided in how to analyze varied aspects of the model text while concurrently helped in the advancement of metalinguistic knowledge. Therefore, it appears that instruction played a pivotal role in broadening our child learners' focus of attention.

Regarding Cycle 2, the children in the LTG seemed to verbalize significantly fewer episodes than their counterparts. As for differences across categories, the groups continued to encounter mostly mechanical, lexical and formal problems, but the LTG also seemed to allocate significantly more attentional resources to such matters as stylistics, paragraphs or coherence and cohesion to a greater extent than in Cycle 1, thus expanding their noticing to a discourse level. Nevertheless, the problems that the TG encountered with content were found to be greater than those of the LTG. Likewise, the self-correction group seemed to show more interest in mechanics than the children in the LTG. In terms of which parts of speech the children were most concerned with, spelling and punctuation were again the primary focus of attention as far as mechanics is concerned; verbs and nouns appertaining to lexis, verb tense and possessive pronouns in the case of form and stylistic issues with regard to discourse. Therefore, in this second cycle we can identify the first differences between groups with respect to Cycle 1, especially in the case of the LTG. The fact that the LTG was better able to attend to a wider array of features tells us something about a plausible influence of a continuous exposure to models on the children's noticing of problematic features. Although at this point we do not have information about what young EFL learners notice after months of treatment with models, it may be the case that the children have developed the ability to anticipate what models would offer them and thus distribute their attention to several aspects while writing the text prior to the reception of the feedback. Besides, the fewer number of episodes discussed by these children may be a consequence of the fatigue produced by repeated exposure to models, or it may simply mean that they did not experience as many problems as their classmates. In any case, it is difficult to offer a

sound explanation for these results and more research is needed to shed light on this issue.

On the other hand, the children's focus on formal aspects when writing a text for the first time in this second cycle is coincidental with the features noticed in Cycle 1, which supports the explanations provided (importance given to grammar in the EFL class and/or influence of the instructional period) for this discrepancy with previous research. However, it could also be argued that this primacy of grammatical forms over meaning among L2 learners is due to procedural task repetition effects. That is, the task was different, but the procedure had been performed several times before. Therefore, in line with Hidalgo and García Mayo (2019), who found that the children in the procedural task repetition condition focused on form to a higher extent than those in the exact task repetition group, it would come as no surprise that our participants focused on grammar as a result of these procedural effects.

In general, these findings further our knowledge of the children's foci of attention during the first stages before and after receiving different treatments. Both cycles were characterized by the noticing of grammar, vocabulary and mechanical problems when composing their texts in response to a visual prompt. In Cycle 2, however, the LTG managed to broaden the scope of their noticing, at the expense of a dominant focus, which translates into a possible effect of the longitudinal treatment.

The objective of research question *1b* was to identify which features the children noticed when comparing their written drafts with model texts or when correcting themselves (Stage 2). Our results revealed that in Cycle 1, the treatment groups noticed not only a significantly higher number of features than the CG, but also more lexical and content-related features than any other category and this difference was also significant with respect to the CG. While the children's high concern about content is contrary to previous research (e.g., Cánovas Guirao et al., 2015; Coyle & Roca de Larios, 2014; García Mayo & Loidi Labandibar, 2017; Hanaoka, 2007; Lázaro-Ibarrola, 2021), the prevailing focus on lexis is consistent with all the findings reported so far (e.g., Cánovas Guirao et al., 2015; Coyle & Roca de Larios, 2014, 2020; Lázaro-Ibarrola, 2021; Luquin & García Mayo, 2020, 2021). For their part, the self-correction

group focused mainly on form and mechanics, the difference of the latter also being significant when compared to the treatment groups' noticing of mechanics.

As can be seen, these findings lay bare that YLs are able to and actually do notice gaps between their written output and the feedback, which supports the commonly held idea that models are valuable for providing learners with a broad range of language input (e.g., Cánovas Guirao et al., 2015; Cánovas Guirao, 2017; Martínez Esteban & Roca de Larios, 2010). Moreover, Hypothesis 1b was supported since, unlike at Stage 1, the overwhelming majority of noticing was lexically driven for the treatment groups at Stage 2 (even when the majority of unsolved problems at Stage 1 were not lexical).

As put forward in the literature, models are usually operationalized through the use of a picture-based descriptive task, which inherently pushes students to describe actions. While performing the task, children strive to find nouns, verbs, adjectives, and adverbs to describe the picture prompt, thus directing most of their attention to lexis and content (Hanaoka & Izumi, 2012; Kang, 2020). In addition, this focus on lexis might also be attributed to the limited processing capacity of the human brain which cannot process content and meaning simultaneously and therefore favors meaning over form, especially at low levels of proficiency (VanPatten, 1990, 2004). During the processing of the model texts, children had to pay attention to meaning and form, notice gaps and holes in their IL, search the model for possible solutions to their initial problems, and use that information to test hypotheses about the TL and formulate new ones. This cognitive overload may have pushed them to turn their attention to semantic and contextual clues when analyzing the model to the detriment of grammatical elements. In like fashion, the slower and unstable development exhibited with formal features agrees with studies on children's L2 oral IL progress which indicates that the acquisition of grammar may be dependent on such factors as age (García Mayo, Lázaro Ibarrola, & Liceras, 2005; Muñoz, 2006), the perceptual saliency of forms or their semantic weight and communicative value (Long, 1991).

The findings observed for the self-correction group indicate that these children attended predominantly to form and mechanics, which concurs with the latest study by Luquin and García Mayo (2021), but not with Luquin and García Mayo (2020), whose participants in the CG spent their time identifying missing information in their writing

and upgrading their texts with new ideational content. This apparent contradiction between the researchers' findings may be related to the small number of episodes generated by the CG in the study in Luquin and García Mayo (2020), which was slightly higher for content, but not significantly different from the remaining categories. It is worth calling attention to this focus on form on the part of the self-correction group since, as suggested in our latest study, it demonstrates that when YLs are provided with no instructions and are left alone with their own drafts, they are able to direct their attention to grammar. This behavior would certainly be an indication of how advantageous writing tasks are to simply engage learners in languaging and draw their attention to formal features. Mechanics, and more specifically spelling, also deserves special mention here considering that the children in the CG attended to mechanics to a significantly higher extent than the treatment groups (and than the LTG at Stage 1 of Cycle 2). This result comes as no surprise given the large amount of spelling and punctuation features the children usually attend to when engaged in collaborative tasks (see Calzada & García Mayo, 2020a, 2020b; Coyle & Roca de Larios, 2014; Lázaro-Ibarrola, 2021; Martínez Esteban & Roca de Larios, 2010). As mentioned above, models have proven effective for drawing students' attention to lexis due (among other issues) to the meaning-focused nature of the task, the type of teaching given to children, or the learners' low proficiency level, which could have affected their dependence on vocabulary to convey basic messages, but data about how children behave when self-correcting their texts is limited. Therefore, it seems that, apart from grammar, spelling and punctuation are also of great concern when children self-edit their drafts.

As for the effect of noticing on subsequent revisions, the TG identified 22% of all the solutions available in the model and the LTG noticed 28% of them, while the learner pairs in the self-correction group verbalized 51 features, out of which 18% had been PFNs. These percentages are significantly low when compared to those obtained with adolescent and adult learners. For instance, the participants in Hanaoka (2006a) reported noticing solutions to 59% of the problematic features, García Mayo and Loidi Labandibar's (2017) teenagers found 63% of the solutions provided in the model; and in the case of Hanaoka and Izumi (2012), their adult learners identified solutions to 69% of their overt problematic features. Notwithstanding the fact that there is a lack of information regarding the relationship between PFNs and FNs within the context of children, it is very probable that a series of factors such as developmental readiness,

proficiency level, working memory or even motivation are the cause of the children's limited awareness and noticing (Cánovas Guirao et al., 2015; Cánovas Guirao, 2017; Manchón, 2014). This is certainly an area that merits further inquiry.

Moving on now to consider the children's noticing at Stage 2 of Cycle 2, we found that the TG seemed to identify a statistically higher number of episodes than the CG and the LTG. We speculate that the groups which had been dealing with their respective forms of feedback for the past five months did not engage in as many episodes due to fatigue caused by task repetition, while the TG arrived fresher at Cycle 2. Nevertheless, the LTG still managed to engage in significantly more LREs and CREs than the CG, maybe due to the nature of the models, which offer richer input than mere self-correction. Anyway, it remains unclear to which degree variability in the amount of LREs and CREs produced is attributed to tiredness, to an effect of sustained exposure to models/self-correction or to any other variable, since this result is not consistent with that obtained at the previous stage (the LTG generated significantly fewer episodes than the other two groups). As for the traceability of the FNs, eight (15%) of the features noticed by the CG at Stage 2 had been PFNs. Therefore, the group behaved similarly in both stages as far as proportion of PFNs over FNs is concerned. As in Cycle 1, the TG found a 22% of all the solvable features, while 24% of the solutions offered by the native text were detected by the children in the LTG. Although statistical significance was not observed between groups, the treatment groups generally exhibited greater noticing of solutions provided by the feedback. With regard to the focus of attention, our results indicate that the CG kept attending mostly to form and mechanics, the TG focused on content, lexis and mechanics, and the learner pairs in the LTG were substantially engaged in formal, lexical and mechanical LREs. From these results it is clear that lexis and content were no longer the only foci when scanning the model text. Form (in the LTG) and mechanics (in both) also gained prominence, while the CG sustained their attention to these two features, exactly as in Cycle 1.

As can be seen, there is not a clear pattern when it comes to comparing the aspects the children paid attention to at the different stages. Nevertheless, these results cast a new light on a potential focus on formal aspects of the model text, as the participants in the model groups seemed to expand their noticing into grammatical elements as well. More specifically, the TG started paying attention to mechanics, whereas the LTG exchanged

content for form. It was the second time dealing with models for the TG and the sixth for the LTG, so it would be interesting to know what further practice with modeling would bring about, leaving the door open to new advantages model texts may offer with greater exposure. The rise in attention to aspects other than lexis confirms a similar tendency reported by Hanaoka (2007b), Martínez and Roca de Larios (2010) and Coyle and Roca de Larios (2014) and implies a valuable role for model texts in fostering noticing among YLs and in helping them diversify their linguistic concerns. As for differences across feedback conditions, the TG was found to attend to both content and lexis to a significantly higher degree than the CG. Contrary to expectations, however, the LTG did not exhibit any difference when compared to the other groups as to their focus of attention. We can observe again that the participants in the LTG distributed their attention to features on equal terms, while the TG continued attending to this second model at a lexical and content-related level to a higher extent than the CG, presumably because the tasks at hand were different.

In essence, at Stage 1, no between-group differences were observed in the total number of CREs and LREs produced in Cycle 1, nor between groups as far as types of features are concerned. The three groups focused their attention mostly on lexis, form and mechanics. At Stage 2 of Cycle 1, we can observe a first difference between the CG and the treatment groups in that the latter (i) noticed significantly more features than the CG when comparing their drafts with the model and (ii) attended to lexis and content to a significantly higher extent than other categories. By contrast, the CG focused mainly on form and mechanics. As a result, these findings suggest a role for models in promoting engagement in a higher number of episodes as well as noticing of lexical and content-related episodes, while self-correction seems to foster children's attention to grammar and mechanical aspects.

At the beginning of Cycle 2, the LTG produced significantly fewer episodes than their counterparts. The children in the CG and the TG continued experiencing mostly mechanical, lexical and formal problems. Nevertheless, apart from these three categories, the LTG also allocated significantly more attentional resources to discourse as opposed to Cycle 1. As for inter-subject differences, the TG attended more to CREs than the LTG, and the CG focused significantly more on mechanics than their classmates in the LTG. At Stage 2, the CG and the LTG engaged in significantly fewer

episodes than the TG, although the LTG still produced significantly more episodes than the CG. Additionally, the three groups showed differences in their focus of attention: On the one hand, while correcting their texts, the CG's focus was entirely on mechanics but also on form over lexis and content. For their part, the children in the TG attended mostly to content and lexis, as they did in Cycle 1, but also to mechanics. Surprisingly, the dyads in the LTG identified formal, mechanical, lexical and content-related features, a wider range of categories than in Cycle 1. Nevertheless, the only statistically significant difference found across groups is that between the TG and the CG in terms of content and lexis.

Finally, research question *1c* closes the analysis of the pair talks by focusing on the comparison of frequency, type and outcome of the episodes between cycles and intra groups.

In relation to frequency, the comparison of the first stages revealed that both the CG and the LTG produced a significantly lower proportion of episodes in Cycle 2, as opposed to Cycle 1. When the comparison stages were contrasted, we found that the LTG engaged in significantly fewer episodes as compared with Cycle 1. This drops on the part of the LTG match with this group's reduction of the total number of episodes produced at Stage 1 and 2 of Cycle 2 as opposed to the TG. As explained above, this decrease may be a consequence of the treatment or simply of a loss of interest, but the why is certainly unclear and therefore further work is required to clear matters up.

Regarding types of features, the LTG produced fewer CREs and grammatical, mechanical and lexical LREs at Stage 1 of Cycle 2, as well as fewer CREs and lexical aspects at stage 2. However, whereas the children only attended to lexis and content at Stage 2 of Cycle 1, they expanded their focus of attention to form and mechanics at this same stage in Cycle 2. We also found a decrease in the number of lexical features noticed by the TG at Stage 1, but these children also happened to observe a higher proportion of mechanical LREs at Stage 2. As mentioned above, it is possible that the more experience the children gain with models, the more attention they pay to aspects other than lexis and content. For their part, the dyads in the self-correction group behaved in the same way as in Cycle 1 as regards their focus of attention. In view of the focus on formal aspects that seems to characterize self-correction, this lack of difference

across cycles confirms that, irrespective of the time spent self-correcting their own texts, the CG kept focusing their attention on form and not so much on meaning.

As for the resolution of episodes, the number of correctly resolved episodes generated by the TG decreased significantly from Cycle 1 to Cycle 2, and the incorrect ones increased. For their part, the CG and the LTG produced a significantly higher number of correctly resolved episodes in Cycle 2, and the LTG also reduced the incorrect ones. The three groups seemed to lower the number of ignored episodes, and no differences were observed for the addressed ones nor for the resolution of the episodes produced by the CG at Stage 2. Consequently, Hypothesis 1c was supported as it was anticipated that the LTG would produce a higher number of (correct) given their previous exposure to models. In a sense, it is accurate to say that the children who had been in contact with models for four months managed to solve their linguistic concerns more successfully than before the treatment. Actually, despite the disparity in the results obtained, this is the most interesting finding to emerge from this analysis. We cannot overlook, however, the fact that the CG also became more accurate in the resolution of episodes, so the possible interference of task-repetition should not be ruled out. On the other hand, we found no differences for the CG at Stage 2, which implies that these children improved their ability to solve their linguistic concerns while writing a text in collaboration but, on the contrary, they were not able to improve the correction of their own mistakes. Although we do not have this information, it is likely that analyzing the outcome of the episodes produced by the LTG at Stage 3, for example, would give us more information about whether this group improved with respect to Stage 3 of Cycle 1. If that was not the case, then task-repetition would be the only reason for the accurate answers given by the learners in the CG and the LTG. Therefore, with the information we have, it is difficult to ascertain whether we are dealing with a task-repetition effect or with a treatment effect. With respect to the TG, the children did considerably worse in Cycle 2, which reveals that interrupting and resuming their contact with models five months later was not the best option for that matter.

Our overall findings on this first set of questions provide further support for the evidence that model texts engage children in a high number of episodes and provide them with rich input of which lexis and content are what they most pay attention to, at least initially. However, the findings also hint at the fact that after a four-month

exposure to models, the children seemed to eventually focus on a wider array of aspects than those observed at the beginning of the study. Self-correction has also proven effective for directing the children's attention to formal aspects, as the participants in the CG directed their attention to grammar, mechanics and punctuation during the whole treatment. Hence, in this sense, we can discard a task-repetition effect on the noticing of LREs and CREs, since the children in the CG did not attend to different aspects of the language. Consequently, the present results are significant in at least two major respects: (i) the use of model texts brings about attention to lexical and content-related aspects but, with enough exposure, they may also be a good option for the children's noticing of formal and discursive aspects of the language. Consequently, it seems that the limited processing capacity of children at these low proficiency levels, which appears to cause greater attention to meaning (VanPatten, 1990, 2004), might be enhanced to the point of widening the 'scope of noticing' (Hanaoka, 2007); and (ii) mere self-correction fosters attention to grammar, spelling and punctuation, which reflects the benefits of writing tasks to engage in languaging.

Last, but not least, the children's pair talk during the four-stage tasks in both cycles also sheds light on the advantages of collaborative writing and feedback analysis with a population which has been underrepresented within the field of SLA. The pairs' overall enthusiasm and eagerness to help each other while jointly exchanging views on their texts and processing the feedback demonstrate that even these inexpert L2 learners were able to combine efforts to create meaning in the L2 and to scrutinize the feedback on the hunt for solutions to their linguistic problems. By placing our participants in proficiency-matched pairs, we definitely obtained a balanced contribution (Storch & Aldosari, 2013; Tedick & Young, 2016), but it is also true that we sacrificed a good performance in the case of the low-low pairs to obtain a high degree of collaboration in all dyads. That is, with fewer linguistic resources to pool, the lower proficiency learners were incapable of fully exploiting the models. This observation coincides with previous research into pair dynamics (Cánovas Guirao, 2017; Storch, 2002, 2005; Storch & Aldosari, 2013; Storch & Wigglesworth, 2010b; Tedick & Young, 2016), which has recognized a series of individual variables such as personality, proficiency level, goals and motivation, that exert an influence on learners' engagement in collaborative writing. Although it was out of the scope of the present dissertation to analyze pair-work

dynamics, the data gathered did reveal that the dyads' level of proficiency had an impact on their performance, as might have been expected.

Research Question 2

How is the children's reported noticing at the composition and comparison stages related to their revised texts?

The integrated analysis of the learners' collaborative talks at stages 3 and 4 and their written output led to several interesting findings. First, with respect to Cycle 1 and in line with previous studies (García Mayo & Loidi Labandibar, 2017; Kang, 2020; Luquin & García Mayo, 2020, 2021), the children in the treatment groups incorporated significantly more features than the CG at both stages 3 and 4, ruling out task-repetition effects. However, this outcome is contrary to that reported by Cánovas Guirao et al. (2015) who found that learners in both groups made a similar number of overall changes in their stories. Apart from this slight discordance, the result confirms that not only did the model groups notice alternative expressions and ideas in the feedback, but they were also able to incorporate them into subsequent revisions. Zooming in on Stage 4 alone, the total number of revisions made by all pairs was substantially low compared to Stage 3 given that the dyads were provided with a different set of pictures and so the likeliness of incorporating the same items was clearly reduced. This finding is consistent with that of Hanaoka (2006a, 2007) and Luquin and García Mayo (2021), who reported a lower incorporation rate on the delayed post-tests. It is also possible that the low proficiency level of the children in combination with their limited working memory and the processing demands of the task hindered the retrieval of noticed items from memory (Cánovas Guirao et al., 2015).

When looking at the types of revisions made by the learner pairs to their drafts, an important distinction is also observable between the CG and the treatment groups in the sense that the former incorporated mostly formal features while the latter made lexical and content-related revisions to their first writings. What is more, when contrasting the three groups at Stage 3, the revised texts in the experimental groups showed signs of lexical and content-related upgrading beyond those of the CG. This overriding focus on lexis and content is coincidental with all previous research on models (e.g., Cánovas Guirao et al., 2015; Coyle & Roca de Larios, 2014; Hanaoka, 2007; Lázaro-Ibarrola, 2021; Luquin & García Mayo, 2020), thus supporting the view that this type of unfocused WCF is beneficial for the incorporation of meaning-oriented elements.

Concerning the performance of the CG, the yields in this investigation do not support previous research including a control condition: Luquin and García Mayo (2020) observed that the changes made by the dyads in the self-correction group were related to lexis, grammar, mechanics and content. Likewise, Kang (2020) found post-test improvements in the CG, and the changes made were also equally distributed across all categories (content, organization, grammar and vocabulary). Luquin and García Mayo (2021) found that, at Stage 3, most of the features incorporated by the CG were lexical and content-related in nature, followed by form and mechanics. This inconsistency may have its origin in the type of noticing happening at the writing stage. That is, the CGs in each study appeared to incorporate into their drafts those features noticed when composing their first text, which may in turn be a reflection of the training session. Stage 4 presents a similar picture, but no significant differences were found across categories or between groups, maybe as a consequence of the few incorporations into the children's second revision. In Luquin and García Mayo (2021), the only study considering the impact of children's noticing on a delayed post-test, it was observed that the features incorporated by the CG at Stage 4 were related to form and mechanics, which is coincidental with our results. However, unlike the present study, the TG incorporated mainly formal and discursive features, but also content-related and lexical items.

With respect to the relationship of these incorporations to both the writing and comparison stages, it was found that most of the features incorporated by the model groups either had not been previously reported (Coyle & Roca de Larios, 2014) or had been identified at the moment of comparison (Cánovas Guirao et al., 2015; Coyle & Roca de Larios, 2014; Lázaro-Ibarrola, 2021; Luquin & García Mayo, 2021), while the majority of the changes introduced by the CG had their origin in the writing stage. These results insinuate that the revisions made by the CG had to do with the problems these children had at the beginning of the task, and not with the potential solutions provided during self-correction. On the contrary, the model groups noticed solutions in the model to both overt and covert problems and incorporated them in their revisions.

To illustrate this, a pair noted as they wrote their first draft that they did not know how to say 'pluma' (feather). They avoided verbalizing this PFN, and instead wrote the following:

‘The clock takes out a mechanic hand and start touching her feet’

While comparing their draft with the model text, a member of the dyad identified a solution to their problem and noted:

(119) Excerpt of conversation:

> CHI47: ‘Stops the clock with her toe’ we put (Keeps on reading). ‘…mechanic hand’ (text), yes. Is… (misses the ‘al’). ‘Feather’! Look!
> CHI48: Feather…?
> CHI47: Feather!

Then at Stage 3, the dyad incorporated this solution:

‘Two minutes later Sara continue sleeping, and the alarm take out a mechanic hand with a feather and starts touching her feet.’ (underline added)

The example above demonstrates that the children were able to recover through the model aspects that they were unsuccessful to encode in their original output (Hanaoka, 2006a). In this sense, models appear to include solutions to both covert and overt problems, thus confirming the results obtained by Hanaoka (2006a), Hanaoka and Izumi (2012) and García Mayo and Loidi Labandibar (2017).

On the other hand, it is also true that while some participants were able to retrieve phrases or chunks of language and incorporate them successfully, as in ‘Martine has sweet dreams’ from ‘Martine is having sweet dreams’, some others extracted these phrases directly from the models as unanalyzed wholes, resulting in inaccurate and/or simplified incorporations, as in ‘It’s Monday morning perfect!’, ‘and Sarah other great school day!’, ‘Another day in the school!’ from the sentence ‘Another great start to the day!’. This finding was also reported by Coyle and Roca de Larios (2014) who, in view of their weakest children’s difficulty in processing the challenging language contained in the model, suggested that feedback should be tailored to learners’ needs. Therefore, it is plausible that the distance between some of the chunks contained in the models and the learners’ L2 knowledge was such that many linguistic features were misunderstood

and unsuccessfully incorporated. Notwithstanding the above, the incorporation of chunks of language is one of the merits of model texts as opposed to EC, which provides targeted feedback, or reformulations, which are limited by its faithfulness to the original writer's intentions. Both feedback techniques fail to provide the writer with a wider array of alternatives above and beyond their own writing (Allwright et al., 1988; Coyle & Roca de Larios, 2014; Martínez Esteban & Roca de Larios, 2010; Yang & Zhang, 2010), while model texts can provide students with a good sample for not only the specific sentence but also the whole discourse. What is more, the use of chunks as a processing strategy is defended by usage-based theories of language learning, which advocate that acquisition is, ultimately, the learning of formulaic sequences (N. Ellis, 2012).

To close Cycle 1 and considering the aforementioned results, we can conclude that Hypothesis 2 was supported, as the learners in the model groups noticed and incorporated mostly lexical features into their subsequent revisions, and most of these changes were traceable to the model or were not previously reported. However, the CG, as predicted, incorporated few of their self-corrections into the first and second revisions.

In Cycle 2, we found the first differences between the three groups in the total number of revisions made, as the model groups incorporated significantly more features than the CG, but the LTG's incorporations were also significantly higher than the TG. These results are likely to be indicative of a positive correlation between the long-term treatment with models and the number of incorporations both in the rewriting stage and on the post-test. Additionally, as in Cycle 1, the treatment groups continued producing a higher number of content-related and lexical revisions than the CG. As for the nature and traceability of the revisions, in the case of the CG, no category stood out from the rest and, as happened in Cycle 1, most of the participants' revisions had originated at the moment of writing their first draft and no FUIs were observed. These results remained the same on the post-test. With regard to the treatment groups, however, it is interesting to note that the most numerous incorporations into their immediate redraft coincide with the categories most attended to during the comparison stage of this same cycle. Namely, the TG incorporated not only CREs and lexical features, but also mechanical elements in its great majority, and once again, most of these incorporations

had either been identified at Stage 2 or had not been previously verbalized. In the case of the participants in the LTG (who had previously attended to lexis, content, mechanics and form), lexis and content excelled from the rest of the categories, but a statistically considerable number of grammatical and mechanical revisions were also made at this point, which means that the LTG not only noticed, but also incorporated a wider range of features than ever before. A similar conclusion was reached by Coyle and Roca de Larios (2014, p. 478), who observed that although lexical revisions were foremost for all their young participants, their results also evidenced the presence of 'a threshold beyond which these less proficient learners might begin to move away from lexis and to diversify their linguistic concerns'.

Another unanticipated finding was that the changes made by the learners in this group originated not only at the comparison stage, but also at the initial stage and that FUIs accounted for more than half of the total incorporations. This finding corroborates the view that the noticing of gaps as a consequence of output production might encourage YLs not only 'to look out for any relevant information available that might help solve the problems in a better way' (Qi & Lapkin, 2001, p. 289), but also to 'use the solutions upon noticing them' (Hanaoka, 2007b, p. 471), which might ultimately result in the children's retention of new L2 forms. As pointed out in previous research (Coyle & Roca de Larios, 2014; Martínez & Roca de Larios, 2010), this retention may also have been promoted in this particular case by the blending of collaborative work and writing implicated in the tasks.

Moving on to the delayed post-test, let us remember that Stage 4 of Cycle 1 was similar to Stage 3, but no statistically significant differences were observed in any of the aspects mentioned up to now. In this case, however, Stage 4 also mirrored Stage 3, but happened to comprise enough information to find significant differences across categories or between groups, which again points to an aftereffect of the long-term treatment with models. As occurred at Stage 3, the children in the CG incorporated significantly fewer items than the treatment groups, the LTG made a higher number of revisions than the TG and also incorporated more vocabulary and more formal features than the former groups. It could therefore be assumed that a long exposure to models could be a major factor, if not the only one, causing the children to incorporate more lexical and formal features than those not benefitting from a long exposure to model

texts. This result is compatible with the theoretical prediction that learning from implicit WCF is more enduring (Kang & Han, 2021). The reason for this may be explained by the fact that the children in the LTG were engaged in deeper levels of processing while comparing their texts to the models. This activity requires reflection and problem-solving which may lead to a cognitive state that moves learning forward, and this effect might have extended to subsequent drafts. When analyzing each group separately, no significant differences were found across categories in the case of the CG, the majority of these incorporations had emerged at Stage 1 and no incorporation was a FUI. As for the TG, the learner pairs remained loyal to what they had focused on from Stage 2 onwards, that is, lexis and mechanics (content was not an option anymore since the story was different), but statistical tests showed a clear supremacy of lexis over any other category. Like the rest of the stages, most of the incorporations had been verbalized at Stage 2 and also a high number of the revisions had not been reported. The main difference found between the treatment groups, observable at both stages 3 and 4, lies in the fact that most of the items incorporated by the TG had either been identified at Stage 2 or had not been previously verbalized, whereas the LTG's revisions originated mainly at both the initial stage and the comparison stage. These results are likely to be related to the fact that, at the moment of writing their first draft in this second cycle, the children commented on aspects they had already encountered in the models used during the past four months and followed through with their noticing at the comparison stage until their incorporation into subsequent drafts. *Examples 120* to *122* below illustrate this:

(120) Excerpt of conversation (S1, C2):

CHI41: The cientific turn into a… 'turn into' or 'turn in a cat'?

CHI42: 'Into', I think it was…

CHI41: And turn into a cat?

CHI42: Yes

(121) Excerpt of conversation (S2, C2):

> CHI41: The ‘turns into a cat’ we put well

Then at Stage 3, the dyad verbalized and incorporated the phrasal verb:

(122) Excerpt of conversation (S3, C2):

> CHI41: And the dog… Yes, and the dog wake up. The sciencist turn into a cat… turn into a cat

As can be seen, the dyad above was able to (i) retrieve the information stored in their long-term memory by recalling the phrasal verb ‘turn into’ from one of the models used during the four-month treatment (see Appendix 15: ‘The witch’), (ii) access their explicit linguistic knowledge, which entails controlled and conscious processing and (iii) incorporate that chunk of language into their first draft. At Stage 2, they encountered this expression in the model, which helped them consolidate it, and at Stage 3, they were better able to retrieve the expression with no signs of hesitation. Even a week later, at Stage 4, the pair continued making use of the phrasal verb:

‘Suddently, the scientist turns into a bat and the dog look terrified.’

Another finding observed at Stage 4, which also emerged at Stage 3, is that a large percentage of the incorporated features had not been noticed or, at least, not reported and surprisingly, some of these incorporations surfaced at Stage 4 and not at Stage 3. Although to a lesser extent, Coyle and Roca de Larios (2014) also reported that their participants in the model group incorporated features that had not been explicitly reported but that later appeared in future texts. The authors attribute this issue to the limitations of note-taking, which is one of the reasons why we decided not to implement this technique. Thus, in view of our similar results concerning FUIs, we have doubts that note-taking might have been the real motive behind these unreported changes. The reasons for these results are not yet entirely understood, but we believe that letting the

gained knowledge settle down for some time may have somehow enabled our young participants to have access to unconsciously learned material at some point. It would be interesting to analyze whether the nature of the episodes contributes to this momentary 'blocking' and how long it would take to free it up (Luquin & García Mayo, 2021). As Cánovas Guirao (2017) points out, in the written output produced by YLs after the provision of WCF we can find evidence which can be indicative of L2 development. Inaccurate occurrences and regressions would seem to indicate that further exposure to feedback is needed for the learning process to progress. Nonetheless, it may also be suggestive of a delayed effect of the WCF, which could become visible in future written output. Consequently, as the Skill Learning Theory stresses, further practice is needed for the consolidation and automatization of that preliminary intake obtained during feedback processing (DeKeyser, 1997). This issue might explain why YLs commonly use the new L2 forms in an irregular or nonlinear way (Bitchener & Storch, 2016). In any case, what is significant is that this variability appears to be indicative of progress. In Ortega's (2012) view, 'any SLA notion of progress has always been nonlinear, gradual, unevenly paced and often proceeding through interim nontargetlike (but developmentally helpful) solutions' (p. 408).

On the question of whether there exist across-cycle differences in the types and frequency of revisions made by the children in the three groups, we only found that the CG incorporated statistically fewer features at Stage 3 of Cycle 2 than at Stage 3 of Cycle 1, whereas the treatment groups underwent no changes. This result tells us that neither a short nor a long treatment with models had an effect on the number of incorporations, but it also indicates that no treatment or, more likely, an excessive use of self-correction might have a detrimental effect. As for variation in the types of categories incorporated, we observed no across-cycle differences in the self-correction condition neither at Stage 3 nor at Stage 4. Statistical analyses showed, however, differences in mechanics for the TG at Stage 3, which largely coincides not only with the rise in mechanical LREs noticed at Stage 2 of Cycle 2, but also with the increase in the incorporation of such features at Stage 3 of Cycle 2. No differences were found between post-tests. Contrary to expectations, no changes were detected for the LTG at Stage 3, but lexis was found to be greater at Stage 4 of Cycle 2, which supports the fact that Stage 4 of Cycle 2 gained strength over that of Cycle 1 for the LTG.

To conclude, apart from analyzing the types and frequency of revisions, we investigated how many of the incorporations were correct and whether there was an effect of models on accuracy in the short and long run. First, we found a similar number of overall correct answers provided by the three groups at both stages 3 and 4 of Cycles 1 (around 75% of all incorporations were accurate) and 2 (around 83%). However, the CG's contribution to the accuracy of the answers was scarce or non-existent. Second, within stages and across groups, the treatment groups always seemed to make more native-like revisions than the self-correction group in Cycle 1, whereas in Cycle 2, the LTG managed to surpass the other two groups. Moreover, the trajectory of the LTG alone revealed a statistically significant improvement in the accuracy of features incorporated into their second writing after five months of treatment. Last, but not least, most of the upgraded revisions made by the model groups were content- and lexis-related (Cánovas Guirao et al., 2015; Luquin & García Mayo, 2020) followed by form, mechanics, discourse and other issues, whereas most of the accurate answers provided by the CG were mostly formal and mechanical in nature.

We cannot forget that the model text failed to get the children to pay attention to and incorporate formal features in the short run as well as to help the dyads integrate some of the chunks or expressions that they had previously noticed. As mentioned in Chapter 4, a number of issues may have constrained our YL's ability to take the most out of models. For example, the children may not have been developmentally ready to take advantage of all the benefits that WCF may offer, because their processing ability is still limited (Pienemann, 1989). Furthermore, according to the limited capacity model (Skehan, 1998), tasks that are excessively demanding absorb learners' attention and memory resources in such a way that students deviate their attention to message content rather than to linguistic form. Pertaining our field of inquiry, Michel et al. (2019) observed that writing and feedback analysis entail complex interactive and recursive cognitive processes. Similarly, Kormos (2012) explained that individual differences in working memory could even determine how L2 learners handle and store information during written production. Finally, the successful processing of WCF is also contingent on the learner's level of proficiency in the L2. More specifically, low proficiency learners may have more difficulty understanding the nature of the gaps they notice from feedback and so benefit less from the feedback (Qi & Lapkin, 2001). They have also been reported to notice more problems with their drafts at the comparison stage than the

higher-level children, which could be interpreted as an aftereffect of going through a higher number of difficulties during text production (Coyle & Roca de Larios, 2014). In addition, less proficient learners may incorporate fewer features than the more advanced ones (Cánovas Guirao et al., 2015); and weaker pairs may engage in poorer collaborative dialogues than stronger dyads (Lapkin, Swain & Smith, 2002). As Manchón (2014) has suggested, any gains from feedback depend on a combination of external and individual factors such as feedback and task types, time of exposure, the linguistic items targeted, learners' cognitive ability, proficiency, attitudes, and beliefs.

On the contrary, these limitations seem to be somewhat alleviated in Cycle 2, and task-procedural effects also seem inexistent, since the LTG appeared to do significantly better than the TG, but also than the CG in terms of noticing and number of (acceptable) incorporations. In addition, as a result of the diversification of their linguistic concerns, this group integrated a broader spectrum of linguistic features in comparison to Cycle 1 and also to their counterparts. All our participants in the model groups found solutions or partial solutions for vocabulary, they upgraded their lexical output by correcting spelling mistakes and incorporated new L2 words and phrases that attracted their attention. These positive results go beyond previous reports and have further strengthened our conviction that the continuous provision of model texts as a WCF technique helped bridge the gap between the children's limitations and the potential learning gains from the feedback. Nevertheless, given that we are dealing with a yet-to-develop population who needs to process unfocused feedback on a meaning-focused task, we strongly side with previous researchers advocating that learners should be trained in how to analyze feedback (Cánovas Guirao, 2017; Cánovas Guirao et al., 2015; Coyle et al., 2018; Qi & Lapkin, 2001; Yang & Zhang, 2010) in order to diversify their noticing. In addition, we suggest that guided instruction should be present throughout the whole process, and not only at the beginning, as evidenced by Cánovas Guirao (2017), whose participants receiving continuous instruction followed more beneficial trajectories than their peers.

In summary, our study provides additional support for the benefits of model texts regarding (acceptable) incorporations in the short term. Accordingly, the revised texts of the learners who had access to the feedback in Cycle 1 showed (i) a significantly higher number of (acceptable) changes than the CG at Stages 3 and 4; (ii) signs of lexical and

content-related upgrading beyond those of the CG, who attended mostly to form; (iii) incorporations that had not been previously reported, which suggests unconscious learning, or had been identified at the moment of comparison, while the majority of the changes introduced by the CG had their origin in the writing stage; and (iv) through mere self-correction, the CG noticed and incorporated many grammatical aspects, which is suggestive of the benefits of writing tasks to engage in languaging. On the other hand, we have also obtained satisfactory results proving that a long exposure to model texts may be effective in alleviating the children's limitations regarding full exploitation of the feedback. Namely, besides the gains obtained in Cycle 1, after four months of treatment with model texts, (i) the LTG managed to make significantly more (acceptable) revisions than any of the groups; (ii) they incorporated a wider range of features than ever before and than their counterparts (the TG also started to focus on aspects different from those noticed in their very first encounter with the feedback); (iii) the traceability of many of the changes stretched to Stage 1, which points to a trail of their previous work with model texts; (iv) for the first time, these three gains were also observed on the post-test; (v) some L2 features emerged at Stage 4 and not at Stage 3, implying a possible delayed effect of the WCF; and (vi) the LTG showed a statistically significant improvement in the accuracy of features incorporated into their second and third writing (or delayed post-test) after five months of treatment. Taken together, we can conclude that uptake of the features noticed at Stages 1 and 2 resulted in the upgrading of the participants' drafts when the items identified were fully or partially incorporated. In addition, it was revealed that the longer the exposure to models, the more diversified attention and, consequently, correct incorporation of features.

7.2. The effects of models on the children's written production

Research Question 3

Do models help improve the written production of primary EFL students in the short and long run?

Evidence of progress in the learners' written output produced in both writing cycles was gathered using a series of parameters consisting of type of clause, lexical diversity, grammatical complexity, fluency and accuracy, analyzed through quantitative measures, as well as adequacy, cohesion, coherence, grammatical accuracy, mechanics and lexical range, evaluated holistically. The main findings observed in each of the aspects analyzed are the following:

The first analysis targeted linguistic acceptability and comprehensibility through the examination of clausal units. The incorporations of feedback into revised texts are typically coded as either accurate or not, as documented in previous studies with adults (see, for example, Yang & Zhang, 2010). This methodological decision completely misses out all the partial gains that children may obtain from WCF, even more so when the ways in which YLs respond to feedback are much less regular and stable. Therefore, it is of utmost importance to use much more nuanced approaches not to overlook students' improvements (Coyle & Roca de Larios, 2014; García Hernández et al., 2017; Villarreal & Lázaro-Ibarrola (submitted)). The texts of this study, produced by learners with limited and insufficient knowledge that in many cases limit the ability to form propositions and articulate them consistently, force the adoption of analytical criteria from the idiosyncratic perspective of the nature of the texts themselves (Torras, 2005). Accordingly, our analysis of the units of language revealed that the model groups reduced the number of pre-clauses on the first post-test. What is more, when the texts of the three groups were compared at this stage, we observed that the treatment groups produced significantly fewer pre-clauses than the CG. This outcome is contrary to that of Coyle and Roca de Larios (2014), who observed that models prompted a higher proportion of pre-clauses in the short term. They attribute this finding to the fact that four of the pairs in this condition were low proficiency. Cánovas Guirao (2017) also found that proficiency had an impact on the types of clausal units the learners produced,

with higher-level dyads writing more proto-clauses and clauses, whereas the written output of the lower-level pairs was characterized by the use of more pre- and proto-clauses both before and after exposure to the model. Both studies examined whether the children's gains were mediated by proficiency, so future studies taking learners' proficiency level into account will need to be undertaken to unveil this issue. In Cycle 2, the TG also produced fewer pre-clauses on the post-test, but not the children in the LTG, whose improvements were instead reflected principally in an increase in clauses.

Regarding long-term results, when draft 1 and draft 6 were compared, we found that the CG wrote significantly fewer clauses after four months of self-edition. The pre-clauses and proto-clauses written by the TG decreased significantly. In the case of the LTG, we found not only significantly fewer pre-clauses, but also fewer proto-clauses. Thus, self-correction may not constitute a good choice to enhance the acceptability of YLs' written texts given the eventual reduction of grammatical clauses. On the other hand, although the children in the TG did not produce more clauses, they managed to reduce the number of pre-clauses and proto-clauses with only two exposures to model texts. All the same, the most optimistic results were obtained by the participants in the LTG, who wrote fewer pre-clauses (like the TG) as well as proto-clauses, but also more clauses between draft 4 and draft 6. This significant increase in clauses within such a short time frame may support the notion that we have been defending so far that the greater the exposure to models, the greater the benefits the children may obtain, at least in certain linguistic aspects. In other words, it may be the case that this last contact with models has triggered the increased production of grammatical clauses. In any case, although further study is warranted given the scant research existing to draw a firm conclusion, these positive findings hint at a possible correlation between models and the linguistic acceptability of the children's written output.

Exposure to the model text within the first and second writing cycles was also found to have positive short-term effects on enhancing the grammatical complexity of the children's third draft, which is visible in the higher number of subordinate clauses produced by the treatment groups on both post-tests. No differences across groups or cycles were found. No differences were observed either when analysing the evolution of lexical diversity indicators from pre- to post-tests in each group, but the results revealed that the LTG showed higher lexical diversity than the other two groups both when they

initiated (draft 4) and finished (draft 6) Cycle 2. The fact that they outperformed the other two groups as soon as Cycle 2 began manifests once again a positive relationship between repeated exposure to models and the variety of words used in the children's texts. As a matter of fact, long-term results revealed that the number of different words used by the learner pairs in the CG and the TG decreased significantly from draft 1 to draft 6. On the contrary, the group benefitting from models for four months showed a significant improvement, visible from their first to their last written text. If we consider other similar studies, we observe mixed results. For example, in their cross-sectional study with children, Villarreal and Lázaro-Ibarrola (submitted) observed that the model group wrote more grammatically and lexically complex texts on the post-test. This finding, however, does not support that of Lázaro-Ibarrola (2021), who did not obtain any text improvements in terms of CAF with a similar sample and after having revised the exact same models as in Villarreal and Lázaro-Ibarrola's (submitted) study. For her part, Cánovas Guirao (2017) reported that all children wrote less grammatical texts than in the first wiring cycle. Given the conflicting results, more textual analyses are needed to determine whether the benefits associated with models lead to gains in lexical diversity and grammatical complexity.

With respect to accuracy, no meaningful short-term findings were observed. The only statistically significant result was for the LTG, who made significantly fewer mistakes from draft 1 to draft 6, and differences were also observed with respect to the CG and the TG in both drafts 4 and 6. The short-term values correlate favorably with Villarreal and Lázaro-Ibarrola (submitted) and Lázaro-Ibarrola (2021), who did not obtain any remarkable gains for any of the groups. The long-term results, however, somewhat mirror the findings obtained in the measures for type of clause, where we observed a significantly higher number of clauses or acceptable units after a long treatment with models. This lack of impact on accuracy in the short run may support the hypotheses of the children's developing metalinguistic ability and the primacy of meaning over form (VanPatten, 2004). Furthermore, it has been suggested that other forms of feedback such as direct EC are more effective to draw children's attention to accuracy, whereas model texts have been praised for their focus on lexis (Chandler, 2003; Coyle & Roca de Larios, 2014). Whatever the reason, it is clear is that one more time these obstacles were saved by the children who had access to several models for the past four months,

as they managed to produce more accurate and thus more acceptable and comprehensible texts.

In line with previous research, fluency appears to be resilient to models. The similarity in the total number of words written before and after exposure to the two feedback techniques insinuates that neither form of feedback has much impact, at least in the short term, on the length of the texts. It seems that the need to convey meaning fosters qualitative changes, but not quantitative ones (Coyle & Roca de Larios, 2014; Lázaro-Ibarrola, 2021; Villarreal & Lázaro-Ibarrola (submitted)). Coyle and Roca de Larios (2014) ascribe this stability of fluency across drafts to the briefness of the texts they were asked to write. In a similar vein, Villarreal and Lázaro-Ibarrola (submitted) attribute this outcome to the task type, which clearly directs the children's attention to content and inhibits more creative attempts. Notwithstanding the above, we did find long-term changes in the number of words produced. When the first and last texts were compared, fluency underwent a pronounced decline in the three groups. That is, all pairs regardless of the feedback condition produced significantly shorter texts in their last writing task as compared to the first one. We interpret this sharp drop in all the children's textual fluency in relation to the end of course. The participants were clearly tired and eager to finish the tasks. Most of them showed a general lack of concentration, motivation and an attitude of indifference which translated into succinct dialogues and shorter texts. In her longitudinal study, Cánovas Guirao (2017) also observed that, as time went on, the children spent shorter time on the tasks and on pair discussions and reduced the length of their texts in comparison to the first cycle. They just wanted to complete the tasks as quickly and with as little effort as possible.

Finally, as for the global analysis of draft quality, the holistic scores showed an across-cycle improvement of the TG's textual cohesion whereas, in line with the quantitative analysis, the LTG enhanced their lexical repertoire from draft 1 to draft 6. When the groups were contrasted, the TG seemed to do significantly better than the CG in terms of coherence on the first post-test, while the LTG obtained a significantly higher score in cohesion than the CG and the TG on the second post-test. Lázaro-Ibarrola (2021) also reported an overall significant short-term improvement when the second draft was evaluated holistically, while Villarreal and Lázaro-Ibarrola (submitted) did not observe any significant differences. The combination of different measurements when analyzing

written production is key to be able to grasp any improvements in draft quality. However, we believe that short-term upgrading may not be as detectable as long-term upgrading when it comes to holistic analysis, since any minor gain obtained in two weeks' time may not be humanly noticeable. In addition, because the arbitrariness of this tool alongside the short range of values provided in the rubric used in this study made it difficult to guarantee a reliable assessment on the children's written texts, the results from such analyses should thus be treated with the utmost caution and only be considered as an add-on analysis to the quantitative assessment.

Taking everything into account, we can conclude that Hypothesis 3 was partially supported: First, in view of the results obtained by Lázaro-Ibarrola (2021) and Villarreal and Lázaro-Ibarrola (submitted) in terms of CAF and global analysis, we hypothesized the non-existence of statistically significant textual improvements triggered by models in the short run. This hypothesis was not supported since we have obtained satisfactory results proving that models are effective to reduce the number of pre-clauses and increase the grammatical complexity of the YL's texts. Second, we speculated that we would find evidence of progress in the written texts, at least with respect to the incorporation of new vocabulary and expressions. We certainly observed that the children were able to produce more acceptable and comprehensible texts, use a wider range of lexical words as well as make fewer mistakes. Therefore, this hypothesis was confirmed. Finally, the CG was also thought to enhance the quality of their texts given the results obtained in one-shot studies. Not only could we not prove this assumption, but we also noticed a downgrading of the learners' final text in terms of type of clause and lexical diversity. In any case, it is important to bear in mind that the negative results obtained for the CG rule out task-repetition effects.

In sum, the short and long-term gains observed through type of clause, CAF and holistic analyses offer further empirical evidence that models can and do help L2 learners upgrade their written output and develop their emerging IL. We can thus confirm that models help improve the overall written production of primary EFL students in both the short and long run. More specifically, in the short-term, models helped the children reduce the number of pre-clauses and proto-clauses as well as increase the grammatical complexity of their texts. After a long exposure to models, the children were able to (i) produce fewer pre-clauses, fewer proto-clauses and more clauses, (ii) use a higher

diversity of lexical words in their texts and (iii) make fewer errors. As was the case with incorporations, practice with model texts appeared to facilitate the feedback processing, decreasing its complexity and enhancing its noticing processes. This increased noticing then resulted in upgrading of their written texts and consequently in L2 development (Cánovas Guirao, 2017). Nonetheless, we believe that practice is not enough when it comes to YLs. If we want children to make the most out of models, we need consciousness raising activities to help younger and weaker learners enhance their meta-awareness of language (Sachs & Polio, 2007).

As a WCF strategy, model texts have proved to constitute a good example of acceptable writing in the L2 as well as a tool to engage children in interaction, 'languaging' and problem solving, thus providing rich opportunities for language learning. When compared to previous research, our findings concerning the effect of model texts on text quality among children cast a new light on the controversy about which aspects of language are most benefitted but also help to reinforce the view that models do have a favorable effect on children's written output. This study also underscores the need to include a wide variety of fine-grained measures not to miss out any minor improvement and the exploration of different tools that corroborate whether small differences do or do not show a potential trend.

Research Question 4

Does input enhancement play a role in the children's noticing and improvement of third person singular possessives?

In order to help the children better exploit the advantages of the model texts, we planned to draw their attention to formal features in three different ways: pre-task instruction, collaborative writing and input enhancement. Research question 4 deals with the third attention-getting tool and is particularly concerned with the extent to which the enhancement of third person possessives can improve the children's knowledge of the targeted linguistic construction. To achieve this aim, we examined the learners' drafts and analyzed the children's development in the use of these pronouns. The following results were observed.

Contrary to expectations, no significant difference was detected across the drafts written by the CG and the TG, but we identified some tendencies. Namely, the self-correction group not only did not show any considerable improvement in the use of third person singular possessives, but became less accurate as the end of the experiment approached. For their part, the TG provided more correct answers in draft 3 of Cycle 1, but their performance turned out to be poorer in Cycle 2. Conversely, the LTG also showed a slight improvement in draft 3, but these gains remained stable throughout Cycle 2, reaching a point at which the children made a correct use of the linguistic target in 100% of the cases on the post-test, showing a total control of third person possessives. Actually, statistically significant differences were observed between drafts 1 and 6. When the three groups were compared, we only observed statistically significant differences between the LTG and the other two groups in draft 6, corroborating the idea that continuous exposure is essential for implicit FonF methods. That is, the LTG achieved greater gains than the TG and the CG on the post-test of Cycle 2. This finding suggests that, although no significant relationships emerged between attention and learning gains of the target linguistic construction in the analysis of each group's development in the short run, significant differences in the long term and across groups were detected, which is evidence of a significant positive correlation between input enhancement and the use of possessive pronouns.

Although it is out of the scope of this dissertation to analyze the children's motivation in this regard, it is interesting to bring into focus some of the comments that emerged from the qualitative data in relation to the enhanced item. As the examples below illustrate, the treatment groups' comments both in the questionnaire (122) and in the focus group interviews (123-125) indicated that the participants noticed the typographically enhanced input and were aware of the target construction of the study.

(123) Questionnaire (LTG):

- *Do you think it might be useful to improve your English?*
- Yes
- *Why?*
- Because I have learned to use *his/her*

(124) Excerpt of conversation (interview – TG)

RES: Have you improved your English? In which aspects?

CHI39: In vocabulary
CHI19: And in grammar
CHI25: And in writing
CHI35: And in the use of 'his' and 'her'

(125) Excerpt of conversation (interview – LTG)

RES: Why did you find the models useful?

CHI48: Because we have learned from our mistakes, and you can learn many more things
CHI57: Also the words… how they are pronounced, how they are spelled…
CHI59: And 'his' and 'her' too

(126) Excerpt of conversation (interview – LTG)

RES: Have you improved your English? In which aspects?

CHI48: In vocabulary

CHI51: Vocabulary

CHI55: Vocabulary

CHI57: And also in the use of 'his' and 'her'… I was very bad at that before…

Therefore, although results were not significant in all directions, Hypothesis 4 was supported: input enhancement had a beneficial effect on the use of third person possessive pronouns. As a matter of fact, we find little agreement in previous research that has analyzed the effects of textual enhancement in reading tasks on grammatical development. To name but a few, Izumi (2002) observed that his adult ESL learners with emerging knowledge in relativization noticed the highlighted relative clauses in the input but failed to show gains in learning. In a more recent study, Lee and Jung (2021) reported the very same result. That is, although their college students did notice the enhanced participle phrases on the reading text, the significant impact of textual enhancement did not extend to the development of L2 grammatical knowledge. In contrast, Lee and Révész (2018) found that textual enhancement not only succeeded in directing their EFL adult learners' attention to the anaphora antecedents, but it also facilitated development in the knowledge of pronominal anaphoric reference. Despite having less proficient learners and a less perceptible target, Lee and Révész's (2018) findings were supported by Chung and Révész (2021). In their longitudinal study, the authors observed that the child EFL learners who were exposed to enhanced third person *-s* morpheme showed small gains, but significantly greater pre-test-post-test improvements as compared to those who engaged in post-reading tasks without textual enhancement.

Let us now turn to a consideration of the factors that might have contributed to the positive results found in our study. As R. Ellis (2016) cautioned, noticing is dependent on a series of factors including the nature of the target form itself, the learner's prior knowledge of it or the level of proficiency. It is difficult to determine which factor, if any, accounts for these relatively optimistic results, or if all these facilitative agents

might have worked together. If we consider the nature of the target form, third person possessives are a salient linguistic form with high communicative value if compared to, for example, third person singular morpheme *-s* (Chung & Révész, 2021), anaphors (Lee & Révész, 2018) or tag questions (Meguro, 2019). For this reason, this feature is more susceptible to being noticed and thus being acquired by language learners from exposure to input alone (Long & Robinson, 1998). Thus, it is possible that the target form in our study was salient enough to draw subjects' attentional resources to form.

In relation to prior knowledge, several researchers (e.g., R. Ellis, 2016; Park, 2004; Winke, 2013) maintain that learners are more likely to allocate their attention to an enhanced form if they have in part acquired it or have some knowledge of it than if it is completely new to them. Although our participants were clearly familiar with third person possessives, we chose this construction in view of the difficulties that many Spanish EFL children seem to experience with distinguishing masculine from feminine possessive pronouns (see Imaz Agirre & García Mayo, 2013, 2018). Actually, this is a feature that does not exist in the children' L1, so its acquisition at these early stages may take time to occur. With reference to this issue, and also connected to the type of target, one of the reasons that Chung and Révész (2021) provide for the relative success of textual enhancement is the children's previous knowledge of the third person *-s* inflection. However, the authors believe that the targeted feature was probably not explicit enough to be perceived in the absence of some predisposition on the part of learners to process the feature. In consequence, the fact that the learners have some previous knowledge of the linguistic form does not guarantee the promotion of L2 learning, so any potential gains would be contingent on the saliency of the enhanced construction.

Notwithstanding the above, R. Ellis (2016) believes that the learner's proficiency level is what most interferes with their allocation of attentional resources and subsequent L2 learning. The theoretical rationale underlying this assumption stems from VanPatten (1990), who proposed that low-proficiency students are not able to process meaning and form simultaneously, leading to greater attention to meaning. This seems to be the case in Lee and Jung (2021), who partly attribute the ineffectiveness of input enhancement in promoting L2 learning to the meaning-focused nature of the task. In other words, the authors explain that the learners focused on understanding the meaning; in consequence,

‘an increase in demands of cognitive resources in comprehending the content of the text might have short-circuited the learning of the L2 grammatical knowledge’ (p. 24). Moreover, the authors added that the participants noticed the enhanced input at the low level but failed to become aware of it at the level of understanding. In our case, dealing with low-proficiency learners might not have worked in our favor as far as conscious noticing of the targeted items is concerned, given this primacy of meaning over form.

However, two variables might have played a positive role: previous knowledge of the content and age. Regarding the former, our children did not face the narrative task from scratch, since they knew about the content of the text from the moment they were in contact with and deciphered the picture prompt at Stage 1. They were therefore better primed to play closer attention to form in their first encounter with the model text. Against this backdrop, Chung and Révész (2021) observed that incorporating textual enhancement into the post-task phase was an effective method to call learners’ attention to form, thus guaranteeing that learners’ attention to meaning was not jeopardized in the during-reading stage of the experiment. For this reason, having access to content beforehand may have alleviated the cognitive overload involved in decoding the meaning of a text. Concerning age, it is true that working with children usually means dealing with low-proficient learners, but it has also been suggested that YLs are more capable of learning languages implicitly than adults, who seem to have lost their implicit mechanisms and therefore need to resort to alternative, explicit mechanisms to learn an L2 (e.g., Bley-Vroman, 1988; DeKeyser, 2000). Thus, in accordance with Chung and Révész (2021), once textual enhancement had drawn our young participants’ attention, implicit mechanisms may have more successfully come into play than in the case of adult learners.

This hypothetical justification is associated with an additional variable that may have contributed to making this attention-drawing device work in the present dissertation. Some researchers (e.g., Cintrón-Valentín, García-Amaya & R. Ellis, 2019; Labrozzi & Villegas, 2020) have claimed that the pedagogical potential of input enhancement could be boosted if combined with other attention-getting tools. As explained in the state-of-the-art section, collaborative writing has been found to be effective in drawing learners’ attention to form while maintaining a primary focus on meaning (R. Ellis, 2016; Swain, 1998, 2000). Consequently, it seems plausible to think that joint work may have acted

as a catalyst for the allocation of the children's attentional resources to the enhanced construction with the subsequent deliberation and sharing of linguistic knowledge in the writing stages. After all, two heads are better than one, but four eyes are also better than two. Therefore, future studies exploring the combination of textual enhancement with other FonF tools seems warranted.

A further reason for the success of textual enhancement in this study might have been the enhancement technique. Indarathne and Kormos (2017) have put forward that underlining might be more effective in generating an isolation effect than other types of input enhancement such as boldfacing (Lee & Révész, 2018) or the use of different color fonts (Lee & Jung, 2021). Some others have used a combination of techniques such as bolding, shadowing, and the use of different fonts and font sizes (Izumi, 2002) or underlining and boldfacing (Chung & Révész, 2021). We decided to follow Indarathne and Kormo's (2017) advice and opted for underlining the chosen linguistic construction. Whether or not the enhancement technique employed here made a difference in the children's accurate use of the linguistic feature remains unknown, and more research is clearly needed to elucidate this question.

Last, but not least, a fundamental characteristic of the present dissertation was the use of a longitudinal design. Our participants in the LTG were exposed to a total of six texts containing underlined third person possessives over a six-month period. Studies addressing the impact of input enhancement over repeated exposure to the enhanced input remain scarce. Among the exceptions we find Chung and Révész's (2021) long-term study undertaken with children. The authors credited the longitudinal design as one of the main reasons for the relative success of textual enhancement and suggest that the next step for YLs to obtain long-term improvements is to retrieve the resulting memory trace repeatedly. Continuous exposure has been alleged to be essential for implicit FonF methods such as input enhancement, and although the learning gains obtained through these implicit methods are typically more enduring, they usually take time to manifest (Mackey & Goo 2007). From this perspective, and in line with Chung and Révész's (2021) results, it is probable that the long-term treatment has increased the odds that the children developed their knowledge of third person possessive pronouns. What is more, the duration of the treatment seemed to be long enough for the gains obtained from input enhancement to flourish.

Summing up, one of the goals of this dissertation was to examine the extent to which input enhancement in model texts can boost learner attention to and subsequent development in the knowledge of third person possessive pronouns. Thereby, we aimed at starting research into the power of visual enhancement in models to have an effect on attention to and development in formal, as opposed to lexical, knowledge. The results revealed that input enhancement succeeded in drawing the children's attention to third person possessives and led to increased gains in the knowledge of the enhanced linguistic feature in the long run. Contrary to expectations, however, we only observed significant differences in the long run and also when the three groups were contrasted, thus corroborating the idea that continuous exposure is essential for implicit FonF methods. We attribute this small but positive long-term result to a combination of factors such as the typographical enhancement technique used to make the input salient, the nature of target linguistic construction, prior knowledge, the age of participants, previous processing of meaning, the use of a longitudinal design and/or collaborative work. Still, it seems necessary to continue exploring the effectiveness of this attention-drawing tool under different conditions and with other pedagogical interventions.

7.3. Learners' attitudes and motivation

Research Question 5

What is the attitude of students toward the three feedback conditions? Does the students' motivation change throughout the treatment?

Apart from the learners' proficiency level, the training session, their working memory and limited processing capacities, the results of the present dissertation are certainly influenced by a series of additional factors that should be considered when providing a rationale for the children's feedback processing. One of the most significant is comprised in the idea of motivation. This last research question aimed to explore the children's attitudes toward the three types of treatment as well as measure their motivation before and after performing the task in both cycles. To this purpose, three questionnaires were administered to the participants. The major findings are summarized below.

Consistent with the literature (Imaz Agirre & García Mayo, 2020; Lázaro-Ibarrola & Villarreal, 2021; Kopinska & Azkarai, 2020; Villarreal & Lázaro-Ibarrola (submitted)), the data from the motivational ratings obtained from the thermometers indicated an overall positive disposition toward the task in the three groups, more positive at post-task (after Stage 3) than at pre-task (before Stage 1) in both cycles, thus supporting Hypothesis 5. As a matter of fact, statistically significant differences between the pre- and post-task were found for the treatment groups in Cycle 1 and for the CG in Cycle 2. The increases in the model groups suggest that the three-stage task was able to raise the learners' initial motivational disposition in the short term, which is coincidental with Lázaro-Ibarrola and Villarreal (2021). On the other hand, the end of the treatment may account for the marginal rise of the CG's motivation in Cycle 2, if we assume that the participants in this group were not precisely motivated, as we will explain below. In a similar vein, the comparison between the first pre-task in Cycle 1 and the last post-task in Cycle 2 revealed significant differences in the case of the self-correction group and the LTG. It is perhaps the sustained activity with writing that caused a drop in motivation observed in these groups. As with the CG in Cycle 2, the increase in motivation exhibited on the last post-test may be explained by the fact that these children, who had been exposed to their respective treatments for six months, became

more motivated as the end of the treatment approached. When motivation ratings were contrasted between groups, we found that no group felt markedly more or less motivated than the others except for the TG in relation to the CG when they wrote the picture story for the second time in Cycle 1. So, for some reason, the TG was motivated enough to make a difference from the CG. It is difficult to explain why the LTG was not as enthusiastic as the TG at this point, as both groups individually showed a significant difference across tasks. What is clear is that models were undoubtedly an element of surprise the first time round, which translated into high levels of motivation.

Notwithstanding all the above, this general upward shift did not reflect reality in all cases, as this high scoring did not match the answers provided in the successive surveys, thus unmasking the true nature of the children's disposition. What the combined information from the three questionnaires truly revealed was that the self-correction group showed a downward fluctuation in motivation, while the treatment groups manifested a more enthusiastic attitude as the study progressed, especially the LTG. The mismatch found in all the answers provided can only have one interpretation: Children usually disclose more sensitive information under anonymous conditions. That is, our participants were required to write down their names in the motivation thermometers, while the rest of the surveys were anonymous. Under the non-anonymous condition, the children probably felt that they had to please the researcher/teacher/parents. Actually, during the interviews, one student asked whether their parents and teacher were going to have access to the interviews, as shown in the example below.

(127) Excerpt of conversation (interview – TG)

RES: What didn't you like about the task?
CHI32: If our parents find out… One thing, wait, are you showing this recording to our parents?
RES: No
CHI32: Ah… And to María? (their teacher)
CHI22: Ah ... Otherwise, I was on my way out

Regarding motives, most of the reasons provided by the learners were positive both in the pre-task and post-task thermometers, leaving negative motives as marginal with only 5 or 6 students, at the most, choosing them. Among the most popular reasons for their high motivation we find having fun, the ease of the task and their eagerness to do an activity in English, but social justifications (working collaboratively) were the most frequently selected motives. Therefore, it seems that collaborative work mostly explains the rise in motivation levels in all groups. Apart from selecting one or more motives from the list, some children took the time to write down their own reasons and it was worthy of note that several students reinforced the value of peer work as a motivating force. As opposed to the scores, this justification surfaced in all three questionnaires, thus corroborating the preferences for collaborative work found among YLs in previous investigations (e.g., Calzada & García Mayo, 2020; Kopinska & Azkarai, 2020; Lázaro-Ibarrola & Villarreal, 2021; Shak & Gardner, 2008; Villarreal & Lázaro-Ibarrola (submitted); Villarreal & Munarriz-Ibarrola, 2021). These results support the value of collaborative work, and model texts, as an appropriate task that engages children in their learning of an L2 (Kopinska & Azkarai, 2020) and also reinforce the idea that motivation is co-constructed (Villarreal & Lázaro-Ibarrola (submitted)).

The next section of the survey was concerned with the learners' opinion on their respective feedback condition, and it consisted of a questionnaire and a focus interview. Taking into consideration the answers provided in the questionnaire and the comments made by the children during the interviews, we can draw some interesting conclusions. Certainly, learners' attitudes towards a given form of feedback are influenced by many factors other than the nature of the technique itself (Loidi Labandibar, 2016). Among other things, we sought to gather information about the use of models as well as self-correction in the children's EFL classes, as previous work has put forward that learners' beliefs are partly determined by the feedback conventions used by their teachers (Kormos, 2012). When the children were asked about their previous experience with the corresponding feedback, the CG reported being familiar with the task, and nearly all of those surveyed from the treatment groups agreed that models had sometimes been used as a feedback technique by their schoolteachers. Conversely, García Mayo and Loidi Labandibar (2017) found that the use of models was totally new for their adolescent participants, and they interpreted this as one of the reasons for the learners' negative attitude towards modelling together with the students' lack of interest in writing, their

lack of motivation to learn English, their low self-efficacy beliefs and the secondary role they think writing plays in L2 development. As we predicted in Hypothesis 5, it is possible that age played a role in our participants' positive disposition, since they are not undergoing emotional changes yet and thus may not be going through the aforementioned issues. Collaboration may also provide a good basis for a positive attitude and, as a result, for a good performance. That is, interaction in writing is still a relatively new concept for teachers, so students who are not used to working jointly may find the task amusing and may feel eager to receive the feedback. Therefore, being an adolescent and working individually may have not worked in their favor.

As for the learners' level of enjoyment, we observed disagreement in the case of the CG. That is, half of the respondents from the CG considered that they had not enjoyed it as much or at all whereas half of them were slightly more optimistic. Models, by contrast, spurred motivation for the task and elicited positive responses, which is in line with what has been reported for adult learners (Hanaoka, 2007; Yang & Zhang, 2010). Thus, as anticipated in Hypothesis 5, the tiresome repetition of self-correction led to an overall pessimistic feeling toward the task, while the use of models in both cycles elicited positive responses from the learners, even more so in the case of the children who had benefited from the models for six months altogether. Precisely, these results obtained for the LTG remind us of those observed in the longitudinal work of Kopinska and Azkarai (2020) on dictogloss tasks, which revealed that the children's disposition was highly positive and seemed to consolidate with time.

In general terms, most of the informants in the self-correction condition acknowledged that this form of feedback might be useful to improve their language skills and learn how to correct their mistakes, whilst some others regarded self-correction as a worthless activity and placed more value on direct corrective feedback. Nevertheless, all of the participants in this group seemed to share the view that correcting their own texts was monotonous, tedious and tiring, to say the least, and because of all these reasons, they made it clear that they would not like their teachers to incorporate this type of feedback in their teaching practices. Leaving correction up to children but also having to repeat this task without serving a real purpose or receiving any type of feedback decreased their motivation toward the task considerably. These results contravene the scores obtained in the motivation thermometer, a mismatch that calls attention to the

significance of data triangulation, as the practice of using multiple sources to analyzing data allows us to obtain answers that reinforce each other, thus enhancing the credibility of the study. Conversely, the treatment groups, especially the LTG, valued the use of models positively, not only in terms of learning, as García Mayo and Loidi Labandibar's (2016) teenagers and Kang's (2020) adults stressed, but also with regard to enjoyment, to the point of wanting their teachers to implement them in the EFL class. Moreover, 'learn', 'fun' and 'English' were the most frequently used words in the questionnaire (see figure 4), followed by 'liked', 'interesting', 'practice', 'good', 'comfortable', 'enjoy', 'well' or 'improve', which represents a fairly visual summary of what models signified to them.

In summary, the tiresome and unrewarding nature of self-correction led to an overall pessimistic feeling toward the task, yet the children acknowledged its effectiveness to learn spelling and vocabulary, among other aspects. In view of this loss of interest, we propose that self-correction should be transformed into a more enjoyable and productive task such as peer correction, as one learner suggested. Oppositely, the use of models elicited positive responses from the participants, even more so in the case of those children who had benefited from this form of feedback for a longer time. Although we cannot overlook the fact that many learners revealed a preference for a more explicit type of error correction, their enjoyment and, overall, their gains in lexis, grammar, writing skills and, more specifically, in the use of third person possessives alongside their enthusiasm to work with their peers certainly make models a powerful pedagogical tool to implement, at least occasionally, in the EFL classroom. All in all, the consideration of individual and contextual factors constitutes an interesting attempt to reach a more comprehensive understanding of the relationship between feedback and L2 learning. Listening to the children's voice has brought us a little closer to knowing whether our tasks and our assessment were effective and appropriate. Given the dynamic and changeable nature of motivation, it is the teacher's responsibility to bring to class tasks which YLs find useful and engaging. In the case of writing, this can represent a real challenge, since writing activities are not as well-received by children as oral communication activities, but we are one step closer to making writing tasks more appealing, effective and interesting.

7.4. Conclusion

This chapter has provided a discussion of our findings in light of each of the hypotheses entertained for the five research questions posed in Chapter 5. Figure 5 illustrates the most notable short- and long-term benefits found for the use of models.

Broadly, the results obtained suggest that models in combination with collaborative work have enabled us to improve many aspects of the children's developing L2 in terms of noticing, writing, learning and incorporation of new L2 features, consolidation of preliminary intake, focus on form, etc. In addition, when looking at the children's performance, it is encouraging to observe that the participants improved their use of third person possessives as well as other formal features, and expanded their vocabulary repertoire. What is more, although performance was not as optimal as that of adult learners given the YLs' limited abilities, we have obtained satisfactory results proving that a long exposure to model texts may be effective in alleviating children's constraints when processing feedback. To facilitate this process and consolidate any potential gains obtained from the feedback, we suggest that repeated exposure should be combined with consciousness raising activities or explicit instruction.

The findings also showed that, in general, the participants were positive about the collaborative writing experience with models. This was not only a consequence of both dealing with models and working in pairs, but also a reason for taking out the most of this feedback technique. The analysis of motivation undertaken here has extended our knowledge of models, strengthened the value of collaborative writing and spurred on the implementation of motivation measures in task-based research.

Figure 5. Benefits of models in our study

The next chapter will provide some concluding remarks as well as pedagogical implications that emerge from this doctoral thesis. In addition, limitations of our study will be presented along with directions for future research.

PART III

CONCLUSIONS AND CONTRIBUTIONS

CHAPTER 8: CONCLUSIONS, CONTRIBUTIONS, IMPLICATIONS AND LINES FOR FUTURE RESEARCH

This final chapter summarizes the main goals of the present doctoral thesis as well as the major conclusions drawn from the results obtained in the data analyses. The main implications of our findings will be presented, and the limitations of our research will be acknowledged. Furthermore, lines for future research on the issues raised will be outlined.

8.1. Conclusions

Certain questions have been raised throughout the literature review that needed (and still need) to be dealt with to fully comprehend the paramount role that collaborative writing and model texts play in the language learning experience of child EFL learners. For example, we addressed the issue of the scarcity of knowledge we have about the impact that child oral interaction has on their written product or about how how primary school-aged children write in collaboration and benefit from WCF in the EFL context. There is also a need to explore alternative WCF techniques, such as models, to shed light on the inconclusive findings on EC, to analyze the benefits of those techniques for YLs, and to examine the characteristics of their interaction when dealing with feedback of this type. Additionally, in order to consider any results as evidence of acquisition and not only of uptake, we highlighted the researchers' calls for longitudinal studies, which would enable the tracking of changes in the YL's written production.

The present study also attempted to explore the role of pedagogical intervention such as pre-task instruction, input enhancement or collaborative reflection to promote the noticing and subsequent (accurate) incorporation of formal features into the children's drafts. Finally, in view of the scarcity of studies about children's attitudes to WCF along with the fact that motivation seems to be fundamental for the students' consolidation of their new knowledge (Bitchener & Storch, 2016), we decided to collect our learners' beliefs and opinions on the specific tasks they performed in order to gain more insights into the motivational processes occurring in the L2 classroom. The research presented here represents an attempt to bridge these gaps in the field by examining the short- and

long-term effects of collaborative writing and model texts on the L2 learning of a group of language learners who have generally been overlooked by researchers of L2 writing.

The major findings of this dissertation are summarized below. The different sections correspond to each of the five research questions posed in chapter 3, thus giving answer to the overall results regarding noticing, oral-written connection, L2 writing, input enhancement, and attitudes and motivation.

(i) With reference to noticing, the use of model texts brought about attention mainly to lexical and content-related aspects but, after prolonged exposure, models were also able to trigger the noticing of formal aspects of the language, and so it was reflected on their drafts, as many of the aspects noticed were fully or partially incorporated. Consequently, it seems that the limited processing capacity of children at these low proficiency levels, which seems to lead to greater attention to meaning (VanPatten, 1990, 2004), might be unblocked to the point of widening their 'scope of noticing' (Hanaoka, 2007) to grammatical aspects. Self-correction fostered attention to grammar, spelling and punctuation, which reflects the benefits of writing tasks to engage in languaging.

(ii) As for the impact of oral interaction on the written mode, uptake of the features noticed at Stages 1 and 2 resulted in the upgrading of the participants' drafts when the items identified were fully or partially incorporated. In addition, we observed that the longer the exposure to models, the more diversified attention and correct incorporation of features.

(iii) Regarding L2 writing, models in the short term led the children to reduce the number of pre-clauses and proto-clauses as well as to increase the grammatical complexity of their texts, which was visible in the higher frequency of subordinate clauses. After a long exposure to models, the children were able to produce fewer proto-clauses and more clauses, use a higher number of different words in their texts and make fewer mistakes.

(iv) With respect to input enhancement, no statistically significant differences were found across the drafts written by the CG and the TG, but some trends were observed. Namely, the CG not only did not improve but their texts became less accurate as the end of the experiment approached. The TG provided more correct answers in draft 3, but

their performance turned out to be poorer in Cycle 2. Conversely, the LTG's enhanced noticing of third person singular possessives in draft 3 remained stable throughout the remaining compositions, reaching a point at which the children used all third person possessive pronouns correctly. Actually, statistically significant differences were observed between drafts 1 and 6. When the three groups were contrasted, we observed statistically significant differences between the LTG and the other two groups in draft 6. That is, the LTG achieved better results than the CG and the TG and this result is likely to be attributable to sustained exposure to the enhanced input.

(v) Concerning attitudes and motivation, the tiresome and repetitive nature of self-correction generated a pessimistic feeling toward the task, but the learners acknowledged its effectiveness to learn spelling and vocabulary, among other aspects. Conversely, the use of models elicited positive responses from the participants, even more so in the case of those children who had benefited from this form of feedback for longer. In general, some learners showed a preference for a more explicit type of error correction, but their enjoyment, improvement in lexis, grammar and writing skills alongside their eagerness to work in collaboration make models a useful tool to implement in the EFL classroom.

8.2. Pedagogical implications

This study has gone some way towards enhancing our understanding of three major spheres from a longitudinal perspective: the language-learning potential of model texts, form-focused intervention and individual attitudes and motivation. Still, our findings should be considered as a way of charting the territory, building on what has been found so far and as a step forward in research on WCF with children. In this section, the pedagogical implications that emerge from the results obtained will be described and some practical suggestions for integrating models in the EFL classroom will be suggested.

8.2.1 The language-learning potential of model texts

The present dissertation provides the first longitudinal assessment of models that includes a large sample of participants, thus allowing statistical analysis that reinforce the results regarding the lasting benefits of modelling on children's L2 development. The findings of this study showcase the need for child EFL learners to be provided with opportunities to write in the L2 and to receive and process feedback on their written output with a view to enhance their L2 knowledge. Likewise, the learners' ability to work in collaboration to detect gaps in their L2 knowledge and then to try to use the new information to upgrade their output shows the value of collaborative writing as a site for learning in the L2 classroom. In EFL contexts, language teachers usually draw on textbooks that, on the pretense of offering communicative language teaching, include a host of formal, lexical and reading exercises arranged in a thematic way. However, writing is frequently consigned to oblivion, since it does not enjoy as much popularity as oral tasks. Hence, it is crucial for educators to be informed about the benefits of writing and feedback processing for L2 learning, as well as of the theoretical implications of writing-to-learn (Cánovas Guirao, 2017; Manchón, 2009).

The use of models in the short run proved useful in helping children to engage in a high number of episodes, expand their lexical repertoires, and incorporate new features into their drafts. After four months of exposure, we observed that the learners diversified their linguistic concerns and integrated a broader spectrum of linguistic features in comparison to Cycle 1 and also in contrast to their counterparts. The implications are clear. Repeated exposure to models seems to facilitate a better use of the feedback by alleviating or unblocking some of the children's limited abilities in terms of working memory, developmental readiness and proficiency level. These optimistic findings go beyond previous reports and have further strengthened our conviction that the continuous provision of model texts as a WCF technique may help bridge the gap between the children's limitations and the potential learning gains from the feedback. If four months of exposure allowed a better use of the models, now the question is which would be the benefits of model texts in the longer run. In any case, considering that we are coping with yet-to-develop learners who need to process unfocused feedback on a meaning-focused task, we strongly recommend that children be trained to analyze

feedback in order to raise their awareness of the widest possible range of aspects. Moreover, we suggest that guided instruction should be present throughout the whole process, and not only at the beginning. In Cánovas Guirao's (2017, p. 258) own words, 'they need to be shown what to look for, how to do it, and then equipped with the metalinguistic knowledge to explain what they find'.

In relation to the abovementioned additional support that children would need and considering the preferences of some children to receive direct written feedback, a second implication is observed. Model texts are not a customized feedback strategy made to fit individual needs. In this sense, it is likely that, for weaker learners, this 'one size fits all' technique might not always be the most suitable solution to correct the children's errors, especially grammar errors (Cánovas Guirao et al., 2015). Consequently, teachers are encouraged to use models in combination with other forms of direct and indirect, focused and unfocused feedback, including EC, reformulation or metalinguistic comments, which may signal more specifically (and offer solutions to) grammatical problems. Model texts are one type of feedback technique, but it is not the only one. Instead of viewing alternative and traditional WCF as a dichotomy, instructors may combine the two approaches for different aims. For example, they can deliver unfocused and/or indirect feedback on learners' first drafts, directing their attention to a great variety of aspects and providing opportunities for self-correction. Direct and/or focused feedback could then be offered to revised drafts for children to correct the remaining errors and produce more accurate texts. Being acquainted with the benefits that different forms of feedback offer and knowing which technique is most appropriate for our learners' needs should be at the forefront of the language teachers' agenda.

To conclude, as previously stated, the mere act of writing has beneficial effects for the learning of an L2, and it is no less true of self-correction. Although most children complained about the unrewarding and tiresome nature of this practice, it is important to bring into contention the language-learning potential of self-evaluation. Even when children are deprived of guidance and must face correction alone, they are capable of allocating their attentional resources to grammatical features and spelling. This behavior is indicative of how helpful writing tasks are to simply engage learners in languaging and draw their attention to formal features. On the other hand, it is true that this technique may turn out to be boresome if abused, but teachers may alternate it with peer

correction, for example, where learners interchange their drafts instead of correcting their own. Therefore, self-correction (or peer correction) should also be a candidate when it comes to selecting the most appropriate type of feedback.

8.2.2 Form-focused intervention

One of the challenges among researchers and teachers is how learner attention can be directed to grammatical forms during communicative tasks. The findings of this dissertation are also of theoretical and pedagogical significance in this regard. By means of the implementation of collaborative writing tasks and the provision of feedback in the form of model texts, it was possible to provide a set of procedures for attracting attention to form while learners were engaged in meaning making. Outside task, pre-task instruction was found useful for flagging children's attention and pointing them to areas of the text that they should reflect on, whereas repeated exposure to WCF enabled the children to begin to move away from lexis and attend to a broader range of features. Within task, building a joint product also raised the dyad's awareness of grammatical features and engaged them in languaging and problem solving. Another pedagogical tool identified as a potential focus-on-form instrument is CF itself, as it is used to induce students' attention to form while completing a task in a personalized and individualized way (R. Ellis, 2005; Van Beuningen, 2010). At this point, we know that model texts are not tailored to the needs of each student, nor are they particularly effective in directing learners' attention to form, at least in the short term, but they in turn proved to be a long-term ideal scenario for the enhancement (and improvement) of targeted linguistic constructions; in this case, of third person possessives.

Increasing the visual salience of third person possessives in model texts for six months led to greater attention allocated to the target feature. Our results, thus, imply that input enhancement, when applied under certain conditions (high saliency of the target form, previous knowledge, underlining, continuous exposure, etc.) can be an effective FonF intervention. An implication of this research is that learning gains on an L2 linguistic construction are very highly related to the attention learners pay to it in the input. Boosting specific grammatical constructions in model texts can direct children's attention to these targeted features and therefore foster L2 development over time. The present doctoral thesis also proves that allowing children to have access to the content

before being exposed to the enhanced linguistic feature can be an effective means of calling their attention to grammar, thus guaranteeing that focus on meaning is not compromised in the during-task phase.

Based on the current findings, it is recommended that EFL instructors make use of sustained implicit and explicit form-focused instruction, as it has proven effective in this particular context. This way, they can provide children with sufficient assistance to achieve the goals that may comprise more difficulty due to their own (or even the feedback's) limitations.

8.2.3 Individual attitudes and motivation

As explained in the literature review, learners' processing of feedback is influenced by both learner and context-related factors. Therefore, the consideration of individual and contextual factors constitutes an interesting attempt to reach a more comprehensive understanding of the relationship between feedback and L2 learning, even more so within the context of YLs. Along these lines, Shak (2006) brought to the fore that childhood education cannot be conceived of without the idea of motivation, especially in classroom contexts where YLs have to deal with arduous and demanding tasks on a daily basis. Actually, if we extrapolate this phenomenon to our study, we find both sides of the coin. On the one hand, we observed that the repetitive and unrewarding practice of self-correction instilled a pessimistic attitude in our learners. On account of this rejection, we propose that teachers should transform self-correction into a more enjoyable and productive task such as peer correction and avoid making extensive use of this technique. On the other hand, the use of models resulted in high motivational dispositions, even more so in the case of those children who had benefited from this form of feedback over an extended period of time. Children should therefore find learning motivating, in such a way that they can develop a sustained level of motivation necessary for long-term achievement (Bitchener & Storch, 2016). Lack of motivation, on the other hand, is an important barrier to academic success, productivity, and wellbeing over time (Legault, Green-Demers & Pelletier, 2006).

Although we cannot ignore the fact that many participants made it clear that they preferred their mistakes to be corrected explicitly, their enjoyment, improvement, and

their eagerness to work hand in hand with their peers certainly make model texts a beneficial pedagogical form of feedback to integrate, at least periodically, into the EFL classroom. Thus, these findings corroborate the value of collaborative work, which EFL instructors should take into account, as well as the usefulness of model texts as a motivating classroom task. As discussed above, we cannot forget that teachers should choose the technique that best suits their learners' needs and the type of task. Accordingly, model texts should complement, and never replace, other forms of feedback.

Considering the dynamic nature of motivation, it is the teacher's responsibility to bring to class tasks which YLs find useful and engaging. As regards writing, this may entail greater difficulty since children do usually not respond as well to writing activities as they do to oral activities given the lack of dynamism, but we are getting closer to making writing tasks more appealing, effective, and engaging over time. In order to know whether our tasks and our assessment are effective and appropriate, teachers need to listen to their learners' needs and beliefs. For example, instructors may make use of entry questionnaires to collect information about students' learning background, objectives, and feedback preferences to help them make feedback choices. Children are far more competent than they are usually given credit for by adults. Therefore, when it comes to teaching and designing new materials, the particular needs, capabilities and perspectives of this young population need to be considered (Pinter, 2011). In Haynes' (2009, p. 3) view, when YLs are offered the opportunity to voice their feelings, this can help to 'change the classroom from a place of instruction into a place where education is possible'.

In line with this issue, we also observed from the range of choices that learners make that what motivates them is dynamism such as modifications from task to task and from learner to learner. This phenomenon should prompt teachers to include variability in their tasks and to measure their learners' motivation as the task progresses. In addition, model texts could be implemented in the classroom to give feedback on different text-types such as essays, reports, letters, narratives, etc. This might be equally effective in CLIL contexts where children are usually asked to write informs or reviews of experiments or other scientific activities. Nevertheless, if the gap between a model and learners' personalized writing is large enough, it may be the case that models are not as

effective with open-ended tasks such as journal, narrative or creative writing (Cánovas Guirao, 2017).

To conclude, in recent years, the use of computer-assisted language learning (CALL) has become almost indispensable in L2 classrooms. The use of CALL in writing tasks involves a series of benefits that can boost learners' motivation to write. For example, children can use technology to communicate with speakers of different languages/cultures. This way, they work together on a written text, collaborating online with others at a distance and using word processing, email, weblogs or wikis to discuss their work (Roblyer, 2016). From a language-learning viewpoint, children would have more time to process input while reading, and edit output while typing, and the fact that their written L2 is visible on the screen may result in greater focus on language form as well as on the feedback received from their partners when communication is synchronous (Cánovas Guirao, 2017).

All in all, it is hoped that our findings will be of value not only to teachers and researchers, but also to the educational authorities that need to make decisions on the best pedagogical practices in the learning of FLs.

8.3. Limitations of the study

We are aware that our work has clearly some shortcomings that need to be acknowledged. Despite this, we believe that these limitations could be a springboard for future work on the topic. First, although no longitudinal study to date has included so many participants, dividing sixty children into three groups and splitting them into pairs resulted in a reduced sample size. Thus, we are conscious that a greater number of dyads would have ensured more representative results. Nevertheless, this was principally a classroom-based study in a real school and, as such, the number of learners who could take part was limited. Notwithstanding the small number of participants in the study, the results are likely to be of more pedagogical importance to instructors of YLs than other more controlled laboratory-based studies of larger subjects.

Second, this research investigates the impact of model texts within and across the two writing cycles that comprised the data collection period but does not offer information that examines the lasting effects of models over a longer period of time (for example, one year later). This data would have allowed us to ascertain whether the children had come to acquire any newly learned material. An additional limitation regarding longitudinal results was that it was not possible to carry out the study right from the start of the academic year due to the busy school schedule. It is true that doubts regarding the effects of regular EFL classes on progress in writing and the development of L2 knowledge would have been raised (Cánovas Guirao, 2017), but this would have allowed us to obtain longer-term results, since the children in the LTG could have been exposed to models for the whole academic year. Another source of weakness in this dissertation was that, although examination of pair dynamics was beyond the scope of this study, we felt that lower-proficiency pairs had a hard time providing support to each other due to their limited knowledge of the L2. Therefore, we consider that a more balanced pairing of children into different proficiency levels is essential to guaranteeing more realistic results.

The study is also limited by the methodological decisions concerning the analysis of the children's writing. The texts object of this study, produced by children at initial stages of L2 learning with limited and insufficient knowledge, forced the researcher to adopt analytical criteria from the perspective of the nature of the texts themselves. Therefore, the low quality and briefness of the written texts implied that the instruments and procedures normally used in research with older learners could not be applied here. Although we decided to replicate other studies (e.g., Cánovas Guirao, 2017; Coyle & Roca de Larios, 2014; Torras, 2005) in the analysis of the children's written output and used the clausal unit measure to capture minor gains across texts, we also wanted to include measures of complexity, accuracy and fluency to yield more detailed information. Even so, the present study may fall short when it comes to providing a full and accurate picture of the children's writing development. Similarly, the arbitrariness of the rubric used in the holistic analysis alongside the short range of values provided in it made it difficult to guarantee a reliable assessment on the children's written output, and therefore the results obtained from such analyses should be treated with caution and only be considered as a complementary analysis to the quantitative assessment.

Continuing with the methodological limitations, this study also failed to fully tailor models to the children's proficiency level. Despite the EFL teachers' consensus that the texts were appropriately adjusted to the learners' level, we found that some chunks were noticed without understanding. For this reason, it seems fundamental to pilot the model texts first and evaluate and adjust them to the level of the participants in the study. This way, we avoid feelings of frustration and get learners to find solutions to the problems they find.

Finally, the study was undertaken under the influence of a couple of external circumstances (such as the proximity of the summer break in Cycle 2 or the lack of a classroom environment to carry out the recordings (see Method section for further details)) that exerted an influence on the children's performance at different times of the data gathering. These factors, which were beyond the researcher's control, are far from reproducing the real conditions under which these learners would typically carry out their daily work in their classrooms. Consequently, it cannot be determined whether the findings reported in the present dissertation would still hold in a real classroom context.

8.4. Future research directions

The exploratory nature of the present dissertation as well as the limitations described above point at new avenues for further research. Firstly, future studies should consider a larger sample of participants to reach more robust conclusions. This dissertation was an experimental study conducted in an EFL setting but future work should be done in classroom settings to have a more ecologically valid scenario of the role that collaborative writing tasks and the use of models play in L2 development. Addressing the issue of the durability of the gains obtained from model texts into future studies would also unravel a series of important questions. On the one hand, it might be clarified to what magnitude children profit from the advantages offered by model texts beyond the data collection period. On the other hand, it might be ascertained whether this form of feedback promotes a more permanent learning of some language features over the others. Longer-term research is thus needed to gauge the permanence of the effects of models and to keep track of the children's progress over a longer period of time.

Taking a more in-depth look at the learners' noticing provides another area worthy of examination. This would involve, for example, analyzing each linguistic category (adjectives, verbs, nouns, noun phrases, verb phrases, agreement, prepositions, etc.) separately. This way we could gather more specific information about which linguistic aspects learners focus on and incorporate the most. In addition, we still ignore why learners uptake data from the models that they have supposedly not noticed and why is it that they only have access to some of these 'unnoticed' features after some time. It would be interesting to analyze whether the nature of the episodes contributes to this momentary 'blocking' and how long it would take to free it up. What is more, a greater focus on the factors that underlie learners' 'scope of noticing' when processing feedback could produce interesting findings that account more for why learners concentrate on some linguistic aspects at the expense of others. This issue is an intriguing one which could be usefully explored in further research.

A key aspect of the present study, that of input enhancement, also merits further investigation. It would be interesting to examine whether the results obtained here would transfer to different language features (e.g., pronouns, prepositions, third person *–s*), populations and contexts or whether its efficacy would change when combined with other pedagogical interventions. Along these lines, further research would also benefit from carrying out an even longer treatment to determine whether continuous instruction, input enhancement, collaborative work and feedback processing could produce even more optimistic results.

Finally, another fruitful idea for future research might be analyzing the children's performance taking into account their L2 proficiency. For instance, although this dissertation did not explore the patterns of pair interaction, the data did suggest that the children's L2 knowledge may have exerted an influence over their ability to help each other. Thus, placing learners in mixed-proficiency pairs, where a more knowledgeable member would assist the weaker learner in producing a written text, could produce different or, at least, interesting findings. Another question related to this issue would be to what extent the relationship between the children's noticing and their revised texts is related to their proficiency levels, or what are the effects of proficiency on the linguistic

acceptability and complexity of the children's drafts. Further longitudinal research that inquired into this matter would be advisable.

As we know, the process of acquiring a new language is an intricate one and different theories and perspectives have been adopted to explain its processes and to inform teaching strategies. This PhD thesis has adopted an interactionist view which underscores the importance of learner collaboration in the learning of an L2. We humbly believe that this study goes one step further towards enhancing our understanding of WCF and that the results reported here will hopefully contribute to advancing the research agenda on feedback for acquisition, provide ideas for FL instructors concerning the provision of WCF and help policymakers make decisions about suitable pedagogical practices.

REFERENCES

Abe, M. (2008). *Exploring the role of model essays in the IELTS writing test: A feedback tool*. (Unpublished master's thesis). University of Queensland, Australia.

Adams, R. (2003). L2 output, reformulation and noticing: Implications for IL development. *Language Teaching Research, 7*, 347–376.

Adams, R. (2007). Do second language learners benefit from interacting with each other? In A. Mackey (Ed.), *Conversational Interaction in Second Language Acquisition* (pp. 29-52). Oxford: Oxford University Press.

Adams, R., & Ross-Feldman, L. (2008). Does writing influence learner attention to form? In D. Belcher & A. Hirvela (Eds.), *The Oral/Literate Connection: Perspectives on L2 Speaking/Writing Connections* (pp. 243-267). Ann Arbor, MI: University of Michigan Press.

Ahmadian, J. (2012). Task repetition in ELT. *ELT Journal, 66*(3), 380-382.

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). Clevedon, UK: Multilingual Matters.

Al Khalil, M. K. (2016). Insights from measurement of task-related motivation. In A. Mackey & E. Marsden (Eds.), *Advancing Methodology and Practice. The IRIS Repository of Instruments for Research into Second Languages* (pp. 243-262). New York, NY & London: Routledge.

Allwright, R. L., Woodley, M. P., & Allwright, J. M. (1988). Investigating Reformulation as a Practical Strategy for the Teaching of Academic Writing. *Applied Linguistics, 9*(3), 236–256.

Ammar, A., & Spada, N. (2006). One size fits all? Recasts, prompts and the acquisition of English possessive determiners. *Studies in Second Language Acquisition, 28*, 543–74.

Azkarai, A., & Kopinska, M. (2020). Young EFL learners and collaborative writing: A study on patterns of interaction, engagement in LREs, and task motivation. *System, 94*, 1-15.

Azkarai, A., & Oliver (2019). Negative feedback on task repetition: ESL vs. EFL child settings. *The Language Learning Journal, 47*(3), 269–280.

Azkarai, A., García Mayo, M. P., & Oliver, R. (2019). The effect of task repetition on the patterns of interaction of ESL children. *International Journal of Applied Linguistics, 171*(1), 90-112.

Barnawi, O. Z. (2010). Promoting noticing through collaborative feedback tasks in EFL college writing classrooms. *International Journal of Teaching and Learning in Higher Education, 22*(2), 209-217.

Basterrechea, M., & Leeser, M. J. (2019). Language-related episodes and learner proficiency during collaborative dialogue in CLIL. *Language Awareness, 28*(2), 97-113.

Batstone, R. (1996). Key Concepts in ELT: Noticing. *ELT Journal, 50*(3), 273-273.

Bitchener, J. (2008). Evidence in support of written corrective feedback. *Journal of Second Language Writing, 17*, 102–118.

Bitchener, J. (2012). A reflection on 'the language learning potential' of written CF. *Journal of Second Language Writing, 21*, 348-363.

Bitchener, J., & Ferris, D. R. (2012). *Written corrective feedback in second language acquisition and writing*. New York: Routledge.

Bitchener, J., & Knoch, U. (2008). The value of written corrective feedback for migrant and international students. *Language Teaching Research Journal, 12*, 409 – 431.

Bitchener, J., & Knoch, U. (2009). The value of a focused approach to written corrective feedback. *ELT Journal, 63*(3), 204-211.

Bitchener, J., & Knoch, U. (2010a). The contribution of written corrective feedback to language development: a ten-month investigation. *Applied Linguistics, 31*(2), 193-214.

Bitchener, J., & Knoch, U. (2010b). Raising the linguistic accuracy level of advanced L2 writers with written corrective feedback. *Journal of Second Language Writing, 19*(4), 207-217.

Bitchener, J., & Storch, N. (2016). *Written Corrective Feedback for L2 Development* (Vol. 96). Bristol, UK: Multilingual Matters.

Bitchener, J., Young, S., & Cameron, D. (2005). The Effect of Different Types of Corrective Feedback on ESL Student Writing. *Journal of Second Language Writing, 14*, 191-205.

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). Rowley, MA: Newbury House.

Bouffard, L. A., & Sarkar, M. (2008). Training 8-year-old French immersion learners in metalinguistic analysis: an innovation in form-focused pedagogy. *Language Awareness, 17*(1), 3-24.

Butler, Y.G. (2015). English language education among young learners in East Asia: A review of current research (2004-2014). *Language Teaching, 48*, 303-342.

Butler, Y. G., & Zeng, W. (2014). Young foreign language learners' interactions during task-based paired assessment. *Language Assessment Quarterly, 11*, 45-75.

Bygate, M. (2001). Effects of task repetition on the structure and control of oral language. In M. Bygate, P. Skehan, & M. Swain (Eds.), *Researching pedagogic tasks: Second Language Learning Teaching and Testing* (pp. 23-48). Harlow: Pearson Education.

Bygate, M. (Ed.). (2018). Learning Language through Task Repetition. Amsterdam: John Benjamins.

Byrnes, H., & Manchón, R. M. (Eds.). (2014). *Task-Based Language Learning – Insights from and for L2 Writing*. Amsterdam: John Benjamins.

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). Amsterdam: John Benjamins.

Calzada, A., & García Mayo, M. P. (2020b). Child learners' reflections about EFL grammar in a collaborative writing task: when form is not at odds with communication. *Language Awareness*. https://doi.org/10.1080/09658416.2020.1751178

Cambridge Young Learners English Assessment. (2018). Cambridge English A2 Flyers. Volume 1. Cambridge English language assessment. Retrieved December 3, 2018, from https://www.cambridgeenglish.org/Images/young-learners-sample-papers-2018-vol1.pdf.

Cameron, L. (2003). Challenges for ELT from the expansion in teaching children. *ELT Journal, 57*, 105-112.

Campillo Ferrer, J. M., López Serrano, S., & Roca de Larios, J. (2012). Spanish children's use of writing strategies when composing texts in English as a foreign language. In M. Torrance, D. Alamargot, & M. Castelló (Eds.), *Learning to write effectively: Current trends in European research* (pp. 191-193). Emerald Group Publishing.

Cánovas Guirao, J. (2017). *The use of written models in the teaching of English in Primary*. (Unpublished doctoral dissertation). Universidad de Murcia, Spain.

Cánovas Guirao, J., Roca de Larios, J., & Coyle, Y. (2015). The use of models as a written feedback technique with young EFL learners. *System, 52*, 63–77.

Carroll, S. (1995) The irrelevance of verbal feedback to language learning. In L. Eubank, L. Selinker, & M. Smith (Eds.), *The Current State of Interlanguage: Studies in Honor of William Rutherford* (pp. 73–88). Amsterdam: John Benjamins.

Chandler, J. (2003). The efficacy of various kinds of error feedback for improvement in the accuracy and fluency of student writing. *Journal of Second Language Writing, 12*(3), 267-296.

Chen, S., Nassaji, H., & Liu, Q. (2016). EFL learners' perceptions and preferences of written corrective feedback: a case study of university students from Mainland China. *Asian-Pacific Journal of Second and Foreign Language Education, 1*, 1–5.

Chung, Y., & Révész, A. (2021). Investigating the effect of textual enhancement in post-reading tasks on grammatical development by child language learners. *Language Teaching Research*. https://doi.org/10.1177/13621688211005068

Cintrón-Valentín, M., García-Amaya L., & Ellis, N. C. (2019). Captioning and grammar learning in the L2 Spanish classroom. *The Language Learning Journal, 47*, 439-459.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Newbury Park, CA: Sage.

Cohen, J. (1992). A power primer. *Psychological Bulletin, 112*(1), 155–159.

Collins, L., & Muñoz, C. (2016). The foreign language classroom: Current perspectives and future considerations. *The Modern Language Journal, 100*, 133-147.

Collins, L., & White, J. (2019). Observing language-related episodes in intact classrooms. In R.M. DeKeyser & G.P. Botana (Eds.), *Doing SLA Research with Implications for the Classroom: Reconciling methodological demands and pedagogical applicability* (pp. 9-30). Amsterdam: John Benjamins.

Copland, F., & Garton, S. (2014). Key themes and future directions in teaching English to young learners: introduction to the Special Issue. *ELT Journal, 68*(3), 223-230.

Corder, S. P. (1967). The significance of learners' errors. *International Review of Applied Linguistics, 5*, 161-169.

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

Coyle, Y., & Roca de Larios, J. (2020). Exploring young learners' engagement with models as a written corrective technique in EFL and CLIL settings. *System, 95*, 1-14.

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.

DeKeyser, R. M. (1997). Beyond explicit rule learning: Automatizing second language morphosyntax. *Studies in Second Language Acquisition, 19*, 195–222.

DeKeyser, R. M. (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: Cambridge University Press.

DeKeyser, R. M. (2000). The robustness of critical period effects in second language acquisition. *Studies in Second Language Acquisition, 22*(4), 499-533.

DeKeyser, R. M. (2007). Skill acquisition theory. In B. VanPatten & J. Williams (Eds.), *Theories in second language acquisition: An introduction* (pp. 97-113). London: Lawrence Erlbaum Associated, Publishers.

DeKeyser, R. M. (2013). Age effects in second language learning: Stepping stones toward better understanding. *Language Learning, 63*(1), 52–67.

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

Doughty, C. J. (2003). Instructed SLA: Constraints, compensation, and enhancement. In C. J. Doughty & M. H. Long (Eds.), *Handbook of second language acquisition* (pp. 256-310). Oxford: Blackwell.

Doughty, C., & Williams, J. (1998). Pedagogical choices in focus on form. In C. J. Doughty & J. Williams (Eds.), *Focus on form in classroom second language acquisition* (pp. 197 – 261). Cambridge: Cambridge University Press.

Dulay, H. C., & Burt, M. K. (1974). Natural sequences in child second language acquisition. *Language Learning, 24*(1), 37-53.

Ellis, G. (2014). '"Young learners": clarifying our terms'. *ELT Journal, 68*(1), 75–8.

Ellis, N. C. (2012). Formulaic language and second language acquisition: Zipf and the phrasal teddy bear. *Annual Review of Applied Linguistics, 32*, 17-44.

Ellis, R. (1994). *The study of second language acquisition*. Oxford: Oxford University Press.

Ellis, R. (1995). Interpretation tasks for grammar teaching. *TESOL Quarterly, 29*, 87-105.

Ellis, R. (1997). *SLA research and language teaching*. Oxford: Oxford University Press.

Ellis, R. (2005). Principles of instructed language learning. *System, 33*(2), 209-224.

Ellis, R. (2009). A typology of written corrective feedback types. *ELT Journal, 63*, 97–107.

Ellis, R. (2010). A framework for investigating oral and written corrective feedback. *Studies in Second Language Acquisition, 32*(2), 335-349.

Ellis, R. (2015). *Understanding second language acquisition*. 2nd edition. Oxford: Oxford University Press.

Ellis, R. (2016). Focus on Form: A critical review. *Language Teaching Research, 20*(3), 405-428.

Ellis, R., Sheen, Y., Murakami, M., & Takashima, H., (2008). The effects of focused and unfocused written corrective feedback in an English as a foreign language context. *System, 36*(3), 353-371.

Enever, J. (Ed.). (2011). *ELLie: Early Language Learning in Europe*. London: British Council.

Enever, J. (2018). *Policy and Politics in Global Primary English*. Oxford: Oxford University Press.

Enever, J., & Lindgren, E. (Eds.). (2017). *Early Language Learning. Complexity and Mixed Methods*. Bristol: Multilingual Matters.

Enever, J., & Moon, J. (2009). New global contexts for teaching primary ELT: Change and challenge. In J. Evener, J. Moon, & U. Raman (Eds.), *Young Learner English Language and Implementation: International Perspectives* (pp. 5-21). Reading: Garnet Education.

Enever, J., & Moon, J. (2010). *A global revolution? Teaching English at primary school*. London, UK: British Council.

Fazio, L. L. (2001). The effect of corrections and commentaries on the journal writing accuracy of minority-and majority-language students. *Journal of Second Language Writing, 10*(4), 235-249.

Fernández Dobao, A. (2014). Vocabulary learning in collaborative tasks: A comparison of pair and small group work. *Language Teaching Research, 18*(4), 497-520.

Ferris, D. (1995). Teaching ESL composition students to become independent self-editors. *TESOL Journal, 4*(4), 18-22.

Ferris, D. (1999). One size does not fit all: Response and revision issues for immigrant student writers. In L. Harklau, K. Losey, & M. Siegal (Eds.), *Generation 1.5 meets college composition* (pp. 143–157). Mahwah, NJ: Lawrence Erlbaum.

Ferris, D. (2002). *Treatment of error in second language student writing*. Michigan: The University of Michigan Press.

Ferris, D. (2004). The 'grammar correction' debate in L2 writing: Where are we, and where do we go from here? (and what do we do in the meantime…?). *Journal of Second Language Writing, 13*, 49-62.

Ferris, D. (2006). Does error feedback help student writers? New evidence on the short- and long-term effects of written error correction. In K. Hyland & F. Hyland (Eds.), *Feedback in second language writing: Contexts and issues* (pp. 81-104). Cambridge: Cambridge University Press.

Ferris, D. (2010). Second language writing research and written corrective feedback in SLA. *Studies in Second Language Acquisition, 32*(2), 181-201.

Ferris, D., & Roberts, B. (2001). Error Feedback in L2 Writing Classes: How Explicit Does It Need to Be? *Journal of Second Language Writing, 10*(3), 161-184.

Fortune, A. (2005). Learners' use of metalanguage in collaborative form-focused L2 output tasks. *Language Awareness, 14*(1), 21-38.

Gagné, N., & Parks, S. (2013). Cooperative learning tasks in a Grade 6 intensive ESL class: Role of scaffolding. *Language Teaching Research, 17*(2), 188-209.

García Hernández, F. J. (2017). *Analysis of the effects of reformulation as a written corrective feedback technique in English with grade six pupils.* (Unpublished doctoral dissertation). Universidad de Murcia, Spain.

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 strategies. In M. P. García Mayo (Ed.), *Learning Foreign Languages in Primary School: Research Insights* (pp. 193-222). Multilingual Matters.

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

García Mayo, M. P. (2018). Child task-based interaction in EFL settings: research and challenges. *International Journal of English Studies, 18*(2), 119-143.

García Mayo, M. P., & Alcón Soler, E. (2013). Negotiated input, output/interaction. In J. Herschensohn & M. Young-Scholten (Eds.), *The Handbook of Second Language Acquisition* (pp. 209-229). Cambridge: Cambridge University Press.

García Mayo, M. P., & Azkarai, A. (2016). EFL task-based interaction. Does task modality impact on language-related episodes? In M. Sato & S. Ballinger (Eds.), *Peer Interaction and Second Language Learning. Pedagogical Potential and Research Agenda* (pp. 241-266). Amsterdam/Philadelphia: John Benjamins.

García Mayo, M. P., & Basterrechea, M. (2013). Dictogloss and the production of the English third person –s by CLIL and mainstream EFL learners: A comparative study. *International Journal of English Studies, 14*(2), 77-98.

García Mayo, M. P., & García Lecumberri, M. L. (Eds.). (2003). *Age and the Acquisition of English as a Foreign Language*. Clevedon: Multilingual Matters.

García Mayo, M. P., & Hidalgo, M. A. (2017). L1 use among young EFL mainstream and CLIL learners in task-based interaction. *System, 67*, 132-145.

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.

García Mayo, M. P., & Imaz Agirre, A. (2017). Child EFL interaction; age, instructional setting and development. In J. Enever & E. Lindgren (Eds.), *Researching the Complexity of Early Language Learning in Instructed Contexts* (pp. 249-268). Bristol: Multilingual Matters.

García Mayo, M. P., & Lázaro Ibarrola, A. (2005). Pronouns revisited: Evidence from the morphosyntactic development of English as an L3. In F. Garrudo Carabias & J. Comesaña Rincón (Eds.), *Estudios de Filología Inglesa en honor de Antonio Garnica* (pp. 217-228). Sevilla: Universidad de Sevilla.

García Mayo, M. P., & Lázaro-Ibarrola, A. (2015). Do children negotiate for meaning in task-based interaction? Evidence from CLIL and EFL settings. *System, 54*, 40-54.

García Mayo, M. P., & Loidi Labandibar, 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.

García Mayo, M. P., Imaz Agirre, A., & Azkarai, A. (2017). Task repetition effect in CAF in EFL child task-based oral interaction. In A. M. Ahmadian & M. P. García Mayo (Eds.), *Recent Perspectives on Task-based Language Learning and Teaching* (pp. 11-28). Berlin: Mouton de Gruyter.

García Mayo, M. P., Lázaro-Ibarrola, A., & Liceras, J. M. (2005). Placeholders in the English Interlanguage of Bilingual (Basque/Spanish) Children. *Language Learning, 55*(3), 445–489.

Gass, S. (1997). *Input, interaction, and the development of second languages*. Mahwah, NI: Lawrence Erlbaum Associates.

Gass, S. M., & Mackey, A. (2006). Input, interaction and output. *AILA Review 19*, 3-17.

Gass, S. M., & Mackey, A. (2007). Input, interaction and output in second language acquisition. In B. VanPatten & J. Williams (Eds.), *Theories in Second Language Acquisition* (pp. 175-199). Mahwah, NJ: Lawrence Erlbaum Associates.

Goldstein, L. M. (2001). For Kyla: What does the research say about responding to ESL writers? In T. Silva & P. Matsuda (Eds.), *On second language writing* (pp. 73-90). Mahwah, NJ: Lawrence Erlbaum.

Goldstein, L. M. (2005). *Teacher written commentary in second language writing classrooms*. Ann Arbor: University of Michigan Press.

Goldstein, L. M. (2006). Feedback and revision in second language writing: Contextual, teacher, and student variables. In K. Hyland & F. Hyland (Eds.), *Feedback in second language writing: Contexts and issues* (pp. 185–205). Cambridge: Cambridge University Press.

Grabe, W. (2009). *Reading in a second language. Moving from theory to practice*. Cambridge: Cambridge University Press.

Gregg, K. R. (1984). Krashen's monitor and Occam's razor. *Applied Linguistics, 5*, 79-100.

Guiraud, P. (1960). *Problèmes et méthodes de la statistique linguistique* [Problems and methods of statistical linguistics]. Dordrecht, The Netherlands: D. Reidel.

Halliwell, S. (1992). *Teaching English in the Primary Classroom*. London: Longman.

Han, Z-H. (2002). Rethinking the role of corrective feedback in communicative language teaching. *RELC Journal, 33*(1), 1-34.

Hanaoka, O. (2006a). Exploring the role of models in promoting noticing in L2 writing. *Jacket Bulletin, 42*, 1–13.

Hanaoka, O. (2006b). Noticing from models and reformulations: A case study of two Japanese EFL learners. *Sophia Linguistica, 54*, 167–192.

Hanaoka, O. (2007). Output, noticing, and learning: An investigation into the role of spontaneous attention to form in a four-stage writing task. *Language Teaching Research, 11*(4), 73–93.

Hanaoka, O., & Izumi, S. (2012). Noticing and uptake: Addressing pre-articulated covert problems in L2 writing. *Journal of Second Language Writing, 21*, 332–347.

Harley, B. (1998). The role of focus-on-form tasks in promoting child L2 acquisition. In C. Doughty & J. Williams (Eds.)*, Focus on form in classroom second language acquisition* (pp. 156–173). Cambridge: Cambridge University Press.

Harley, B., & Swain, M. (1984). The interlanguage of immersion students and its implication for second language teaching. In A. Davies, C. Criper, & A.P.R. Howatt (Eds.), *Interlanguage* (pp. 291-311). Edinburgh, Scotland: Edinburgh University Press.

Haynes, J. (2009). Listening to the Voice of Child in Education. In S. Gibson and J. Haynes (Eds.), *Perspectives on Participation and Inclusion: Engaging Education* (pp. 27–41). London: Continuum.

Hedgcock, J., & Lefkowitz, N. (1992). Collaborative oral/aural revision in foreign language writing instruction. *Journal of Second Language Writing, 1*(3), 255–276.

Hidalgo, M. A. (2017). *The oral production and negotiation of meaning of Spanish EFL children in task-supported interaction: The role of age and task repetition*. (Unpublished doctoral dissertation). University of the Basque Country, Spain.

Hidalgo, M. A., & García Mayo, M. P. (2019). The influence of task repetition type on young EFL learners' attention to form. *Language Teaching Research, 25*(4), 565-586.

Hidalgo, M. A., & 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.

Horst, M., White, J., & Bell, P. (2010). First and second language knowledge in the language classroom. *International Journal of Bilingualism, 14*(3), 331–349.

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

Huang, B. H. (2015). A synthesis of empirical research on the linguistic outcomes of early foreign language instruction. *International Journal of Multilingualism, 13*, 257-273.

Hyland, F. (1998). The impact of teacher written feedback on individual writers. *Journal of Second Language Writing, 7*, 255–286.

Hyland, F. (2011). The language learning potential of form-focused feedback on writing: Students' and teachers' perceptions. In R. Manchón (Ed.), *Learning- to-write and writing-to-learn in an additional language* (pp. 159–179). Amsterdam: John Benjamins.

IBM Corp. Released 2020. IBM SPSS Statistics for Windows, Version 27.0. Armonk, NY: 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.

Imaz Aguirre, A., & García Mayo, M. P. (2018). Proficiency and transfer effects in the acquisition of gender agreement by L2 and L3 English learners. In J. Cho, M. Iverson, T. Juddy, T. Leal & E. Shimanskaya (Eds.), *Meaning and Structure in Second Language Acquisition. In honor of Roumyana Slabakova* (pp. 203-227). Amsterdam: John Benjamins (Studies on Bilingualism Series).

Imaz Agirre, A., & García Mayo, M. P. (2019). Task modality and pair formation method: Their impact on patterns of interaction and LREs among EFL primary school children. *System, 80*, 165-175.

Imaz Agirre, A., & García Mayo, M. P. (2020). The impact of agency in pair formation on the degree of participation in young learners' collaborative dialogue. In C. Lambert & R. Oliver (Eds.), *Using Tasks in Second Language Teaching: Practice in Diverse Contexts* (pp. 306-323). Multilingual Matters.

Indrarathne, B., & Kormos, J. (2017). Attentional processing of input in explicit and implicit conditions: An eye-tracking study. *Studies in Second Language Acquisition, 39*, 401–430.

Indrarathne, B., & Kormos, J. (2018). The role of working memory in processing L2 input: Insights from eye-tracking. *Bilingualism: Language and Cognition, 21*, 355–374.

Izumi, S. (2002). Output, input enhancement, and the noticing hypothesis: An experimental study of ESL relativization. *Studies in Second Language Acquisition, 24*, 541–577.

Izumi, S. (2003). Comprehension and production processes in second language learning: In search of the psycholinguistic rationale of the Output Hypothesis. *Applied Linguistics, 24*(2), 168-196.

Izumi, S. (2013) Noticing and L2 development: Theoretical, empirical, and pedagogical issues. In J. M. Bergleithner; S. N. Frota, & J. Kei, (Eds.), *Noticing and second language acquisition: Studies in honor of Richard Schmidt* (pp. 25-38). University of Hawai'i at Manoa. National Foreign language Resource Center.

JASP Team (2021). JASP (Version 0.16.2)[Computer software].

Johnson, K. (1988). Mistake correction. *ELT Journal, 42*, 89-96.

Kang, E.Y. (2020). Using model texts as a form of feedback in L2 writing. *System, 89*, 1-11.

Kang, E.Y., & Han, Z. (2015). The efficacy of written corrective feedback in improving L2 written accuracy: a meta-analysis. *Modern Language Journal, 99*, 1–18.

Kang, E. Y., & Han, Z. (2021). Written corrective feedback: Short-term and long-term effects on language learning. In R. Manchón, & C. Polio (Eds.), *The Routledge Handbook of Second Language Acquisition and Writing* (pp. 213-225). New York: Routledge.

Kearney, E., & Ahn, S-Y. (2014). Preschool world language learners' engagement with language: what are the possibilities? *Language Awareness, 23*, 319-333.

Kellogg, R. T. (2008). Training writing skills: A cognitive development perspective. *Journal of Writing Research, 1*, 1–26.

Kim, Y. (2008). The contribution of collaborative and individual tasks to the acquisition of L2 vocabulary. *Modern Language Journal, 92*(1), 114−130.

Kim, Y., & McDonough, K. (2011). Using pre-task modelling to encourage collaborative learning opportunities. *Language Teaching Research, 15*(2), 183-199.

Klein, W. (1986). *Second Language Acquisition*. Cambridge University 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.

Kormos, J. (2012). The role of individual differences in L2 writing. *Journal of Second Language Writing, 21*(4), 390-403.

Kormos, J., & Trebits, A. (2012). The role of task complexity, modality and aptitude in narrative task performance. *Language Learning, 62*(2), 439–472.

Krashen, S. (1977). The monitor model for adult second language performance. In M. Burt, H. Dulay, & M. Finocchiaro (Eds), *Viewpoints on English as a second Language* (pp. 152-161). New York: Regents.

Krashen, S. (1981). *Second language acquisition and second language learning*. Oxford: Pergamon Press.

Krashen, S. (1982). *Principles and practice in second language acquisition*. Oxford: Pergamon Press.

Krashen, S. (1985). *The Input Hypothesis: Issues and Implications*. London: Long-man.

Labrozzi, R. M., & Villegas, A. (2020). The effects of dynamic visual enhancement and working memory on the recall of Spanish vocabulary. *Hispania, 103*, 245-258.

Lalande, J. F. (1982). Reducing composition errors: An experiment. *The Modern Language Journal, 66*(2), 140-149.

Lantolf, J., & Aljaafreh, A. (1995). Second language learning in the Zone of Proximal Development: *A revolutionary experience. International Journal of Education Research, 23*, 619-632.

Lantolf, J. P., & Thorne, S. L. (2006). *Sociocultural Theory and the Genesis of Second Language Development*. Oxford: Oxford University Press.

Lapkin, S., Swain, M., & Smith, M. (2002). Reformulation and the learning of French pronominal verbs in a Canadian French immersion context. *The Modern Language Journal, 86*, 485-507.

Larson-Hall, J. (2010). *A guide to doing statistics in second language research using SPSS.* New York: Routledge. Chicago.

Lázaro-Ibarrola, A. (2002). *La adquisición de la morfosintaxis del inglés por niños bilingües (euskera/castellano): Una perspectiva minimalista* [The acquisition of English morphosyntax by bilingual (Basque-Spanish) children: A minimalist perspective]. Ph.D. dissertation, Universidad del País Vasco/Euskal Herriko Unibertsitatea (UPV/EHU), Vitoria-Gasteiz, Spain.

Lázaro-Ibarrola, A. (2009). Reformulation and self-correction: testing the validity of correction strategies in the classroom. *RESLA, 22*, 189-215.

Lázaro-Ibarrola, A. (2021). Model texts in collaborative and individual writing among EFL children: Noticing, incorporations and draft quality. *International Review of Applied Linguistics in Language Teaching.* https://doi.org/10.1515/iral-2020-0160

Lázaro-Ibarrola, A., & Azpilicueta Martínez, R. (2015). Investigating negotiation of meaning in EFL children with very low levels of proficiency. *International Journal of English Studies, 15*, 1-21.

Lázaro-Ibarrola, A., & García Mayo, M. P. (2012). L1 use and morphosyntactic development in the oral production of EFL learners in a CLIL context. *International Review of Applied Linguistics in Language Teaching, 50*(2), 135-160.

Lázaro-Ibarrola, A., & Hidalgo, M. A. (2017a). Benefits and limitations of conversational interactions among young learners of English in a CLIL context. In M. P. García Mayo (Ed.), *Learning Foreign Languages in Primary School. Research Insights* (pp.86-102). Bristol: Multilingual Matters.

Lázaro-Ibarrola, A., & Hidalgo, M. A. (2017b). Procedural repetition in task-based interaction among young EFL learners. *ITL: International Journal of Applied Linguistics, 168*, 183–202.

Lázaro-Ibarrola, A., & Villarreal, I. (2021). Are EFL Writers Motivated or Demotivated by Model Texts and Task Repetition?: Evidence from Young Collaborative Writers. *International Journal of English Studies, 21*(2), 29–55. https://doi.org/10.6018/ijes.466401

Lee, I. (2004). Error correction in L2 secondary writing classrooms: The case of Hong Kong. *Journal of Second Language Writing, 13*, 285–312.

Lee, I. (2009). Ten mismatches between teachers' beliefs and written feedback practices. *ELT Journal, 63*(1), 12-22.

Lee, M., & Jung, J. (2021). Effects of textual enhancement and task manipulation on L2 learners' attentional processes and grammatical knowledge development: A mixed methods study. *Language Teaching Research*. https://doi.org/10.1177/13621688211034640

Lee, M., & Révész, A. (2018). Promoting grammatical development through captions and textual enhancement in multimodal input-based tasks. *Language Learning from Multimodal Input, 42*(3), 625-651.

Lee, S. K. (2007) Effects of textual enhancement and topic familiarity on Korean EFL students' reading comprehension and learning of passive form. *Language Learning, 57*, 87–118.

Leeser, M. J. (2004). Learner proficiency and focus on form during collaborative dialogue. *Language Teaching Research, 8*(1), 55-81.

Legault, L., Green-Demers, I., & Pelletier, L. G. (2006). Why do high school students lack motivation in the classroom? Toward an understanding of academic amotivation and the role of social support. *Journal of Educational Psychology, 98*, 567-582.

Leont'ev, A. N. (1978). *Activity, consciousness, and personality*. Englewood Cliffs, NJ: Prentice Hall.

Leont'ev, A. N. (1981). The problem of activity in psychology. In J. V. Wertsch (Ed.), *The concept of activity in Soviet psychology* (pp. 37-71). Armonk, NY: Sharpe.

Leow, R. P. (2015). *Explicit learning in the L2 classroom. A Student-Centred Approach*. New York. Routledge.

Levenston, E. A. (1978). Error analysis of free comparison: The theory and the practice. *Indian Journal of Applied Linguistics, 4*(1), 1–11.

Lewis, A. (1992). Group child interviews as a research tool. *British Educational Research Journal, 18*(4), 413-421.

Li, S., & Vuono, A. (2019). Twenty-five years of research on oral and written corrective feedback in System. *System, 84*, 93-109.

Lightbown, P. M. (1983). Acquiring English L2 in Quebec classrooms. In S. Felix & H. Wode (Eds.), *Language development at the crossroads* (pp. 101–120) Tubingen: Gunter Narr.

Lo, J., & Hyland, F. (2007). Enhancing students' engagement and motivation in writing: The case of primary students in Hong Kong. *Journal of Second Language Writing, 16*, 219 – 237.

Loewen, S. & Sato, M. (2018). Interaction and instructed second language acquisition. *Language Teaching, 51*, 285-329.

Loidi Labandibar, U. (2016). *The use of models as written corrective feedback in spontaneous EFL writing*. (Unpublished master's thesis). University of the Basque Country, Spain.

Long, M. H. (1983). Native speaker/non-native speaker conversation and the negotiation of comprehensible input. *Applied Linguistics, 4*, 126-141.

Long, M. H. (1988). Instructed interlanguage development. In L. M. Beebe (Ed.), *Issues in second language acquisition: Multiple perspectives* (pp. 115-141). Cambridge, MA: Newbury House/Harper and Row.

Long, M. H. (1991). Focus on form: A design feature in language teaching methodology. *Foreign language research in cross-cultural perspective, 2*(1), 39-52.

Long, M. H. (1996). The role of the linguistic environment in second language acquisition. In T. K. Bhatia & W. C. Ritchie (Eds.), *Handbook of Language Acquisition* (pp. 413-468). New York, NY: Academic Press.

Long, M. H. (2000). Focus on form in task-based language teaching. In R. Lambert & E. Shohamy (Eds.), *Language policy and pedagogy. Essays in honor of A. Ronald Walton* (pp. 179-192). Philadelphia: John Benjamins.

Long, M. H., & Robinson, P. (1998). Focus on form: Theory, research, and practice. In C. Doughty & J. Williams (Eds.), *Focus on form in second language acquisition* (pp. 15-41). Cambridge: Cambridge University Press.

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

Luquin, M., & García Mayo, M. P. (2020). Collaborative writing and feedback: An exploratory study of the potential of models in primary EFL students' writing performance. *Language Teaching for Young Learners, 2*(1), 73-100.

Luquin, M., & García Mayo, M. P. (2021). Exploring the use of models as a written corrective feedback technique among EFL children. *System, 98*(1), 1-13. https://doi.org/10.1016/j.system.2021.102465

Lyster, R. (1998). Negotiation of form, recasts, and explicit correction in relation to error types and learner repair in immersion classrooms. *Language Learning, 48*(2), 183-218.

Lyster, R. (2004). Different effects of prompts and recasts in form-focused instruction. *Studies in Second Language Acquisition, 26*(3), 399-432.

Lyster, R. (2007). *Learning and Teaching Languages through Content*. Amsterdam: John Benjamins.

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

Lyster, R., & Ranta, L. (1997). Corrective feedback and learner uptake: Negotiation of form in communicative classrooms. *Studies in Second Language Acquisition, 19*, 37-66.

Mackey, A. (1999). Input, interaction and second language development: An empirical study of question formation in ESL. *Studies in Second Language Acquisition, 21*(4), 557-587.

Mackey, A. (Ed.) (2007). *Conversational Interaction in Second Language Acquisition. A Collection of Empirical Studies*. Oxford: Oxford University Press.

Mackey, A. (2020). *Interaction, feedback and task research in second language learning: Methods and design*. Cambridge University Press.

Mackey, A., Abbuhl, R., & Gass, S. M. (2012). Interactionist approaches. In S. M. Gass & A. Mackey (Eds.), *Handbook of Second Language Acquisition* (pp. 7-23). New York: Routledge.

Mackey, A., & Goo, J. (2007). Interaction research in SLA: A meta-analysis and research synthesis. In A. Mackey (Ed.), *Conversational Interaction in Second Language Acquisition* (pp. 407-472). Oxford: Oxford University Press.

Mackey, A., & Oliver, R. (2002). Interactional feedback and children's L2 development. *System, 30*, 459-477.

Mackey, A., & Silver, R. E. (2005). Interactional tasks and English L2 learning by immigrant children in Singapore. *System, 33*(2), 239-260.

Makino, T. (1979). English morpheme acquisition order of Japanese secondary school students. Abstract of Ph.D. dissertation. *TESOL Quarterly, 13*, 428.

Manchón, R. M. (2009). Broadening the perspective of L2 writing scholarship: The contribution of research on foreign language writing. In R. M. Manchón (Ed.), *Writing in foreign language contexts. Learning, teaching, and research* (pp. 1-19). Clevedon: Multilingual Matters.

Manchón, R. M. (Ed.) (2011). *Learning-to-Write and Writing-to-Learn in an Additional Language*. Amsterdam: John Benjamins.

Manchón, R. M. (2014). The internal dimensions of tasks. In H. Byrnes & R. M. Manchón (Eds.), *Task-based language learning insights from and for L2 writing* (pp. 27-53). Amsterdam: John Benjamins.

Manchón, R.M. (2018). Past and future research agendas on writing strategies: Conceptualizations, inquiry methods, and research findings. *Studies in Second Language Learning and Teaching, 8*(2), 247-267.

Manchón, R. M., & Roca de Larios, J. (2007). On the temporal nature of planning in L1 and L2 composing: A study of foreign language writers. *Language Learning, 57*(4), 549-593.

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). New York: Routledge.

Martínez Esteban, N., & Roca de Larios, J. (2010). The use of models as a form of written feedback to secondary school pupils of English. *International Journal of English Studies, 10*, 143–170.

McCutchen, D. (2011). From novice to expert: Implications of language skills and writing-relevant knowledge for memory during the development of writing skill. *Journal of Writing Research, 3*, 51–68.

McLaughlin, B. (1987). *Theories of second-language learning*. London: Edward Arnold.

Meguro, Y. (2019). Textual enhancement, grammar learning, reading comprehension, and tag questions. *Language Teaching Research, 23*(1), 58-77.

Michel, M., Kormos, J., Brunfaut, T., & Ratajczak, M. (2019). The role of working memory in young second language learners' written performances. *Journal of Second Language Writing, 45*, 31-45.

Montealegre Ramón, F. (2020). Exploring the learning potential of models with secondary school EFL learners. *Spanish Journal of Applied Linguistics, 32*(1), 155-184.

Muñoz, C. (Ed.). (2006). *Age and the Rate of Foreign Language Learning*. Clevedon: Multilingual Matters.

Muñoz, C. (2007). Age-related differences and second language learning practice. In R. DeKeyser (Ed.), *Practice in a Second Language. Perspectives from Applied Linguistics and Cognitive Psychology* (pp. 229-255). Cambridge: Cambridge University Press.

Muñoz, C. (2014). The interaction of L1 timing, onset age and input in second/foreign language acquisition. *Linguistic Approaches to Bilingualism, 4*, 368-372.

Murakami, A., & Alexopoulu, T. (2016). L1 influence on the acquisition order of English grammar morphemes. A learner corpus. *Studies in Second Language Acquisition, 38*, 365-401.

Murphy, L., & Roca de Larios, J. (2010). Feedback in second language writing: An introduction. *International Journal of English Studies, 10*(2), 1-15.

Murphy, V. (2014) *Second language learning in the early school years. Trends and contexts*. Oxford: Oxford University Press.

Nassaji, H. (1999). Towards integrating form-focused instruction and communicative interaction in the second language classroom: some pedagogical possibilities. *Canadian Modern Language Review, 55*, 385–402.

Ndlovu, K., & Geva, E. (2008). Writing abilities in first and second language learners with and without reading disabilities. In J. Kormos & E. H. Kontra (Eds.), *Language learners with special needs: An international perspective* (pp. 36–62). Clevedon: Multilingual Matters.

Nicholas, H., & Lightbown, P.M. (2008). Defining child second language acquisition, defining roles for L2 instruction. In J. Philp, R. Oliver, & A. Mackey (Eds.), *Second language acquisition and the younger learner: Child's play?* (pp. 27–52). Amsterdam/Philadelphia: John Benjamins.

Nikolov, M. (1999). "Why do you learn English?" "Because the teacher is short": A study of Hungarian children's foreign language learning motivation. *Language Teaching Research, 3*, 33-65.

Nikolov, M., & Mihaljevic-Djigunović, J. (2011). All shades of every color: An overview of early teaching and learning of foreign languages. *Annual Review of Applied Linguistics, 31*, 95-119.

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

Norman, G. (2010). Likert scales, levels of measurement and the "laws" of statistics. *Advances in Health Sciences Education, 15*, 625-632.

Oliver, R. (1995). Negative feedback in child NS-NNS conversation. *Studies in Second Language Acquisition, 17*, 459–481.

Oliver, R. (1998). Negotiation of meaning in child interactions. *Modern Language Journal, 82*(3), 372-386.

Oliver, R. (2000). Age difference in negotiation and feedback in classroom and pair work. *Language Learning, 50*(1), 119–151.

Oliver, R. (2002). The patterns of negotiation for meaning in child interaction. *The Modern Language Journal, 86*, 97-111.

Oliver, R. (2009). How young is too young? Investigating negotiation of meaning and 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). New York, NY: Routledge.

Oliver, R., & Azkarai, A. (2017). Review of child second language acquisition (SLA): Examining theories and research. *Annual Review of Applied Linguistics, 37*, 62-76.

Oliver, R., & Grote, E. (2010). The provision and uptake of different types of recasts in child and adult ESL learners: What is the role of age and context? *Australian Review of Applied Linguistics, 33*(3), 26.1–26.22.

Oliver, R., & Mackey, A. (2003). Interactional context and feedback in child ESL classrooms. *The Modern Language Journal, 87*(4), 519-533.

Oliver, R., Nguyen, B., & Sato, M. (2017). Child ISLA. In S. Loewen & M. Sato (Eds.), *The Routledge handbook of instructed second language acquisition* (pp. 468–487). New York, NY: Routledge.

Ortega, L. (2009a). Studying writing across EFL contexts: Looking back and moving forward. In R. Manchón (Ed.), *Writing in Foreign Language Contexts: Learning, Teaching and Research* (pp. 232-255). Clevedon: Multilingual Matters.

Ortega, L. (2009b) *Understanding Second Language Acquisition*. New York. Routledge.

Ortega, L. (2012). Epilogue: Exploring L2 writing-SLA interfaces. *Journal of Second Language Writing, 21*, 404–415.

Park , E. S. (2004). Constraints of implicit focus on form: Insights from a study of input enhancement. *Teachers College, Columbia University Working Papers in TESOL & Applied Linguistics, 4*(2), 1-30.

Pei-sui Luk, X., & 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.

Philp, J., Oliver, R., & Mackey, A. (2006). The impact of planning time on children's task-based interactions. *System, 34*(4), 547-565.

Philp, J., Oliver, R. & Mackey, A. (Eds.). (2008). *Second Language Acquisition and the Younger Learner: Child's Play?* Amsterdam: John Benjamins.

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. (1994). Research on negotiation: What does it reveal about second language learning conditions, processes, and outcomes? *Language Learning, 44*, 493-527.

Pica, T. (2002). Subject matter content: How does it assist the interactional and linguistic needs of classroom language learners? *The Modern Language Journal, 86*, 1-19.

Pica, T. (2013). From input, output and comprehension to negotiation, evidence, and attention. An overview of theory and research on learner interaction and SLA. In M. P. García Mayo, J. Gutierrez Mangado, & M. Martínez Adrián (Eds.), *Contemporary Approaches to Second Language Acquisition* (pp. 49-70). Amsterdam: John Benjamins.

Pienemann, M. (1989). Is language teachable? Psycholinguistic experiments and hypotheses. *Applied linguistics, 10*(1), 52-79.

Pienemann, M. (2007). Processability theory. In B. VanPatten & J. Williams (Eds.), *Theories in second language acquisition: An introduction* (pp. 137-154). Mahwah, N. J.: Lawrence Erlbaum.

Pinter, A. (2005) Task repetition with 10-year-old children. In C. Edwards & J. Willis (Eds.), *Teachers Exploring Tasks in English Language Teaching* (pp. 113-126). Basingstoke: Palgrave Macmillan.

Pinter, A. (2006). Verbal evidence of task related strategies: Child versus adult interactions. *System, 34*(4), 615–630.

Pinter, A. (2007). Some benefits of peer-peer interaction: 10-year-old children practicing with a communication task. *Language Teaching Research, 11*, 189-207.

Pinter, A. (2011). *Children Learning Second Languages*. London: Palgrave Macmillan.

Pinter, A. (2014). Child participant roles in applied linguistics research. *Applied Linguistics, 35*, 168-183.

Pinter, A. (2017). *Teaching Young Language Learners* (Second Edition). Oxford: Oxford University Press.

Polio, C. (2012). The relevance of second language acquisition theory to the written error correction debate. *Journal of Second Language Writing, 21*, 375 – 389.

Polio, C., Fleck, C., & Leder, N. (1998). "If only I had more time": ESL learners' changes in linguistic accuracy on essay revisions. *Journal of Second Language Writing, 7*(1), 43-68.

Polio, C., & Williams, J. (2009). Teaching and testing writing. In M. H. Long & C. Doughty (Eds.), *The handbook of language teaching* (pp. 486–517). Oxford: Wiley-Blackwell.

Qi, D. S., & Lapkin, S. (2001). Exploring the role of noticing in a three-stage second language writing task. *Journal of Second Language Writing, 10*, 277–303.

Reinders, H. (2009). Learner uptake and acquisition in three grammar-oriented production activities. *Language Teaching Research, 13*(2), 201–222.

Robinson, P. (1995). Attention, memory and the ‘‘noticing’’ hypothesis. *Language Learning, 45*, 283–331.

Robinson, P. (2003). Attention and memory during SLA. In C. Doughty & M. H. Long (Eds.), *Handbook of second language acquisition* (pp. 630-678). Oxford: Blackwell.

Roblyer, M. D. (2016). *Integrating Educational Technology into Teaching* [Chapter 2: Theory into Practice – Foundations for Effective Technology Integration]. Upper Saddle River, N.J.: Prentice Hall.

Roca de Larios, J. (1999). *Cognitive processes in L1 and L2 writing*: *a cross-sectional study*. (Unpublished doctoral dissertation). Universidad de Murcia, Spain.

Sachs, R., & Polio, C. (2007). Learners’ uses of two type of written feedback on a L2 writing revision task. *Studies in Second Language Acquisition, 29*(1), 67-100.

Sajavaara, K. (1981). The nature of first language transfer: English as L2 in a foreign language setting. Paper presented at the first European-North American Workshop on Second Language Acquisition research, Lake Arrowhead, Calif.

Sample, E., & Michel M. (2014). An Exploratory Study into Trade-off Effects of Complexity, Accuracy, and Fluency on Young Learners’ Oral Task Repetition. *TESL Canada Journal, 31*(8), 23-46.

Santos, M., López-Serrano, S., & Manchón, R. M. (2010). The differential effect of two types of direct written corrective feedback on noticing and uptake: Reformulation vs. error correction. *International Journal of English Studies, 10*(1), 131-154.

Schmidt, R. (1990). The role of consciousness in second language learning. *Applied Linguistics, 11*(2), 129-158.

Schmidt, R. (1994). Deconstructing consciousness in search of useful definitions for applied linguistics. *AILA Review, 11*, 11–26.

Schmidt, R. (1995). *Attention and Awareness in Foreign Language Learning* (Vol. 9). Honolulu: University of Hawai’i Press.

Schmidt, R. (2001). Attention. In P. Robinson (Ed.), *Cognition and second language instruction* (pp. 3–32). Cambridge: Cambridge University Press.

Schmidt, R., & Frota, S. (1986). Developing basic conversational ability in a second language: A case study of an adult learner of Portuguese. In R. Day (Ed.), *Talking to learn: Conversation in second language acquisition* (pp. 237–326). Rowley, MA: Newbury House.

Segalowitz, N. (2000). Automaticity and attentional skill in fluent performance. In H. Riggenbach (Ed.), *Perspectives on fluency* (pp. 200-219). Ann Arbor, MI: University of Michigan Press.

Segalowitz, N., & Lightbown, P. (1999). Psycholinguistic approaches to SLA. *Annual Review of Applied Linguistics, 19*, 43-63.

Shak, J. (2006). Children using dictogloss to focus on form. *Reflections on English Language Teaching, 5*, 47-62.

Shak, J., & Gardner, S. (2008). Young learner perspectives on four focus-on-form tasks. *Language Teaching Research, 12*(3), 387-408.

Sharwood-Smith, M. (1974). Teaching written English: problems and principles. *E.T. Forum, 12*(3), 1-15.

Sharwood-Smith, M. (1991). Speaking to many minds: On the relevance of different types of language information for the L2 learner. *Second Language Research, 7*(2), 118–132.

Sheen, Y. (2004). Corrective feedback and learner uptake in communicative classrooms across instructional settings. *Language Teaching Research, 8*(3), 263-300.

Sheen, Y. (2006). Exploring the relationship between characteristics of recasts and learner uptake. *Language Teaching Research, 10*(4), 361–392.

Sheen, Y. (2007). The effect of focused written corrective feedback and language aptitude on ESL learners' acquisition of articles. *TESOL Quarterly, 41*(2), 255-283.

Sheen, Y. (2010). Differential effects of oral and written corrective feedback in the ESL classroom. *Studies in Second Language Acquisition, 32*(2), 201-234.

Shook, D. (1999). What foreign language reading recalls reveal about the input-to-intake phenomenon. *Applied Language Learning, 10*, 39–76.

Simard, D., Guénette, D., & Bergeron, A. (2015). L2 learners' interpretation and understanding of written corrective feedback: insights from their metalinguistic reflections. *Language Awareness, 24*(3), 233-254.

Simon, E. (2010). Second language acquisition and the younger learner: Child's play? *Journal of Child Language, 37*(4), 945–951.

Skehan, P. (1998). *A cognitive approach to language learning*. Oxford University Press.

Spada, N., & Lightbown, P. M. (1999). Instruction, first language influence, and developmental readiness in second language acquisition. *The Modern Language Journal, 83*(1), 1-22.

Storch, N. (2001). How collaborative is pair work? ESL tertiary students composing in pairs. *Language Teaching Research, 5*(1), 29-53.

Storch, N. (2002). Patterns of interaction in ESL pair work. *Language Learning, 5*, 119–158.

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

Storch, N. (2013). *Collaborative writing in L2 classrooms*. Bristol, UK: 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). Berlin: De Gruyter.

Storch, N. (2018). Written corrective feedback from sociocultural theoretical perspectives: A research agenda. *Language Teaching, 51*(2), 262-277.

Storch, N. (2019). Collaborative writing. *Language Teaching, 52*, 40–59.

Storch, N. (2021). Collaborative writing: Promoting languaging among language learners. In M. P. García Mayo (Ed.), *Working Collaboratively in Second/Foreign Language Learning* (pp. 13–34). Berlin: De Gruyter Mouton.

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.

Storch, N., & Aldosari, A. (2013). Pairing learners in pair work activity. *Language Teaching Research, 17*, 31–48.

Storch, N., & Wigglesworth, G. (2003). Is there a role for the use of the L1 in an L2 setting? *TESOL Quarterly, 32*(4), 760−770.

Storch, N., & Wigglesworth, G. (2010a). Learners' processing, uptake, and retention of corrective feedback on writing: Case studies. *Studies in Second Language Acquisition, 32*(2), 303–334.

Storch, N., & G. Wigglesworth, G. (2010b). 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: Oxford University Press.

Swain, M. (1985). Communicative competence: Some roles of comprehensible input and comprehensible output in its development. In S. Gass & C. Madden (Eds.), *Input in second language acquisition* (pp. 235-253). Rowley, MA: Newbury House.

Swain, M. (1991). French immersion and its offshoots: Getting two for one. In B. Freed (Ed.), *Foreign language acquisition: Research and the classroom* (pp. 91-103). Lexington, MA: Heath.

Swain, M. (1993). The output hypothesis: Just speaking and writing aren't enough. *The Canadian Modem Language Review, 50*, 158-164.

Swain, M. (1995). Three functions of output in second language learning. In G. Cook & B. Seidhofer (Eds.), *Principles and practice in applied linguistics* (pp. 125-144). Oxford: 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: Cambridge University Press.

Swain, M. (2000). The output hypotheses and beyond: Mediating acquisition through collaborative dialogue. In J. Lantolf (Ed.), *Sociocultural theory and second language learning* (pp. 97–114). Oxford: Oxford University Press.

Swain, M. (2001). Integrating language and content teaching through collaborative tasks. *Canadian Modern Language Review, 58*(1), 44-63.

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). Mahwah, NJ: Lawrence Erlbaum Associates.

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). London: Continuum.

Swain, M. (2010). 'Talking-it-through': languaging as a source of learning. In R. Batstone (Ed.), *Sociocognitive perspectives on language use/learning* (pp. 112-130). Oxford: Oxford University Press.

Swain, M., & Lapkin, S. (1995). Problems in Output and the Cognitive Processes They Generate: A Step Towards Second Language Learning. *Applied Linguistics, 16*(3), 371–39.

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). London, UK: Pearson Education.

Swain, M., & Lapkin, S. (2002). Talking-it-through': Two French immersion learners' responses to reformulation. *International Journal of Educational Research, 37*, 285-304.

Tedick, D. J., & Young, A. I. (2016). Fifth grade two-way immersion students' responses to form-focused instruction. *Applied Linguistics, 37*(6), 784–807.

Tocalli-Beller, A., & Swain. M. (2005). Reformulation: The cognitive conflict and L2 learning it generates. *International Journal of Applied Linguistics, 15*, 5 28.

Tomlin, R. S., & Villa, V. (1994). Attention in cognitive science and second language acquisition. *Studies in Second Language Acquisition, 16*(2), 183-203.

Torras, R. (2005). Procesos psicolingüísticos implicados en la adquisición del inglés en el contexto de la enseñanza primaria [Psycholinguistic processes involved in learning English in a primary school context]. *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 the Rate of Foreign Language Learning* (p. 156-182). Bristol: Multilingual Matters.

Truscott, J. (1996). The case against grammar correction in L2 writing classes. *Language Learning, 46*(2), 327-369.

Truscott, J. (1998). Noticing in second language acquisition: a critical review. *Second Language Research, 14*(2), 103–135.

Truscott, J. (2009). Arguments and appearances: A response to Chandler. *Journal of Second Language Writing, 19*(1), 59-60.

Truscott, J., & Sharwood-Smith, M. (2011). Input, intake, and consciousness. *Studies in Second Language Acquisition, 33*(4), 497-528.

Uggen, M. S. (2012). Reinvestigating the noticing function of output. *Language Learning, 62*(2), 506-540.

Van Beuningen, C. (2010). Corrective Feedback in L2 Writing: Theoretical Perspectives, Empirical Insights, and Future Directions. *International Journal of English Studies, 10*(2), 1-27.

Van Beuningen, C. G., De Jong, N. H., & Kuiken, F. (2012). Evidence on the effectiveness of comprehensive error correction in second language writing. *Language Learning, 62*(1), 1-41.

Van den Branden, K. (1997). Effects of negotiation of on language learners' output. *Language Learning*, 47, 589-636.

VanPatten, B. (1990). Attending to form and content in the input. *Studies in Second Language Acquisition, 12*, 287-301.

VanPatten, B. (2004). Input processing in SLA. In B. VanPatten (Ed.), *Processing instruction: Theory, research and commentary* (pp. 5–32). Mahwah, NJ: Erlbaum.

Vasylets, O., Gilabert, R. & Manchón, R. M. (2019). Differential contribution of oral and written modes to lexical, syntactic and propositional complexity in L2 performance in instructed contexts. *Instructed Second Language Acquisition, 3*(2), 206-227.

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. 89-116). Berlin, Boston: De Gruyter Mouton. https://doi.org/10.1515/9781501511318-005

Villarreal, I. & Lázaro-Ibarrola, A. (submitted). Models in collaborative writing among CLIL learners in Primary School: Linguistic outcomes and motivation matters.

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

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

Waller, L., & Papi, M. (2017). Motivation and feedback: How implicit theories of intelligence predict L2 writers' motivation and feedback orientation. *Journal of Second Language Writing, 35*, 54-65.

Wigglesworth, G., & Storch, N. (2012). What role for collaboration in writing and writing feedback. *Journal of Second Language Writing, 21*(4), 364-374.

Wilcox, R. (2001). *Fundamentals of modern statistical methods: Substantially improving power and accuracy*. New York: Springer.

Williams, J. (1999). Learner-generated attention to form. *Language Learning, 49*(4), 583- 625.

Williams, J. (2001). The effectiveness of spontaneous attention to form. *System, 29*, 325- 340.

Williams, J. (2012). The potential role(s) of writing in second language development. *Journal of Second Language Writing, 21*, 321–331.

Winke, P. (2013). The effects of input enhancement on grammar learning and comprehension. *Studies in Second Language Acquisition, 35*(2), 323-352.

Wong, W. (2003). Textual enhancement and simplified input: Effects on L2 comprehension and acquisition of non-meaningful grammatical form. *Applied Language Learning, 13*, 17–45.

Yamashita, T. (2021). Corrective feedback in computer-mediated collaborative writing and revision contributions. *Language Learning & Technology, 25*(2), 75–93.

Yang, L., & Zhang, L. (2010). Exploring the role of reformulations and a model text in EFL students' writing performance. *Language Teaching Research, 14*, 464–484.

Yarrow, F., & Topping, K. J. (2001). Collaborative writing: The effects of metacognitive prompting and structured peer interaction. *British Journal Of Educational Psychology, 71*(2), 261.

Zhang, S. (1995). Re-examining the affective advantage of peer feedback in the ESL writing class. *Journal of Second Language Writing, 4*, 209-222.

APPENDIXES

Appendix 1: Background questionnaire

	Información general
1.	**Nombre:**
2.	**¿Cuántos años tienes?**
3.	**¿Eres NIÑO o NIÑA?**
4.	**Curso:**
5.	**¿Cuál es la primera lengua que aprendiste en casa?** 1. Euskara 2. Castellano 3. Las dos 4. Otra (¿cuál?)_____
6.	**¿En qué idioma hablas con las siguientes personas?** Con tu madre _____ Con tu padre _____ Con tus hermanos/as _____ Con tus amigos _____
7.	**Hablo….** Castellano 1. Nada 2. Poco 3. Bien 4. Perfecto Euskara 1. Nada 2. Poco 3. Bien 4. Perfecto Inglés 1. Nada 2. Poco 3. Bien 4. Perfecto Otro (¿cuál?) _____ 1. Nada 2. Poco 3. Bien 4. Perfecto Otro (¿cuál?) _____ 1. Nada 2. Poco 3. Bien 4. Perfecto
8.	**¿A qué edad EMPEZASTE a estudiar inglés? _____ años**
9.	**¿ Además de la asignatura de inglés, ¿has estudiado alguna otra asignatura en inglés?** ☐ **No** ☐ **Sí, ¿cuál?_____**
10.	**Además de en el colegio, ¿has recibido alguna vez clases particulares o extraescolares de inglés?** ☐ **SI** ☐ **NO** **(Si la respuesta es SÍ)…** **¿Cuántas horas semanales? _____ horas**

11. **¿Recibes ahora clases particulares o extraescolares de inglés?**

☐ **SI** ☐ **NO**

(Si la respuesta es SÍ)…

¿Cuántas horas semanales? _____ horas

12. **¿Has estado alguna vez en un país de habla inglesa?**

☐ **SI** ☐ **NO**

(Si la respuesta es SÍ)…

¿Cuántas veces y por cuánto tiempo? _____

13. **¿Cuántas horas semanalas usas el inglés fuera de clase para hacer las siguientes cosas?:**

	0h	**1-2h**	**3-4h**	**5-6h**	**7-8h**
Leer libros					
Ver la tele, películas, series, …					
Escuchar musica					
Surfear por internet (youtube, …)					
Jugar videojuegos					
Hablar con amigos o familiares					
Otro: ….					

Adapted from the questionnaire of the project 'Interaction and written production. The potential of collaborative writing in the learning of English as a foreign language by primary school learners' (FFI2016-74950-P).

English version

General information	
1.	**Name:**
2.	**How old are you?**
3.	**¿Are you a boy or a girl?**
4.	**Academic course:**
5.	**Mother tongue (first language learned at tome)** 1. Basque 2. Spanish 3. Both 4. Other (Specify)_____
6.	**¿ Languages you speak at home:** With your mother _____ With your father _____ With your siblings _____ With your friends _____ With other people (specify) _____
7.	**I speak….** Basque 1. A little 2. Some 3. Well 4. Near-native Spanish 1. A little 2. Some 3. Well 4. Near-native English 1. A little 2. Some 3. Well 4. Near-native Other (specify) 1. A little 2. Some 3. Well 4. Near-native Other (specify) 1. A little 2. Some 3. Well 4. Near-native
8.	**How old were you when you FIRST started learning English? _____ years**
9.	**Besides your English classes, have you studied any other subject in English?** ☐ **No** ☐ **Yes (specify)_____**
10.	**Besides studying at school, have you taken any English courses in a private school and/or private tutoring?** ☐ **SI** ☐ **NO** **(If YES)…** **How many hours per week? _____ hours**

11. **¿Are you now taking private or extracurricular English classes?**

☐ **YES** ☐ **NO**

(If YES)…
How many hours per week? _____ hours

12. **Have you ever travelled to an English-speaking country to study English?**

☐ **YES** ☐ **NO**

13. **¿How many hours per week do you spend using English outside class to do the activities listed below?:**

	0h	**1-2h**	**3-4h**	**5-6h**	**7-8h**
Read books					
Watch TV, films, series, …					
Listen to music					
Surf the internet (youtube, …)					
Play videogames					
Talk to friends and/or familiy					
Other: ….					

Appendix 2: Informed consent

HOJA DE INFORMACIÓN Y CONSENTIMIENTO INFORMADO

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 investigación. Es necesario que lea detenidamente la siguiente información y que pregunte si tiene alguna duda al respecto.

CONTACTO:
Investigadora principal: María del Pilar García Mayo
Dirección: Paseo de la Universidad 5
Centro: Facultad de Letras
Teléfono: 945 013036
Correo electrónico: mariapilar.garciamayo@ehu.eus

DATOS RELATIVOS AL PROYECTO:

- **Título del proyecto**: Interacción y producción escrita. El potencial de la escritura colaborativa en el aprendizaje de inglés como lengua extranjera por parte del alumnado de Educación Primaria
- **Financiado por el Ministerio de Economía y Competitividad**
- 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 una tarea oral y otra de producción escrita.
    - Ambas tareas serán tareas colaborativas que se completarán en parejas.
    - En la tarea oral, los participantes describirán y narrarán en parejas una historia reflejada en unas viñetas donde tendrán que consensuar las descripciones de cada imagen.
    - En la tarea de producción escrita, los participantes completarán una tarea similar a la anterior pero esta vez, además de consensuar la historia, deberán decidir sobre una producción escrita basada en la narración.
    - Datos personales anónimos: los datos personales serán tratados de forma totalmente anónima, así como los resultados de todas las pruebas.
- Número de intervenciones: la recogida de datos se realizará en cuatro sesiones de 15 minutos de duración aproximada cada una en enero y junio en el Colegio San Viator
- Descripción del procedimiento: En cada sesión el/la participante completará una tarea colaborativa junto a un compañero/a con un límite de tiempo de aproximadamente 15 minutos. Las tareas tienen como objeto obtener

información sobre los beneficios de las tareas colaborativas y la recepción de feedback 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 a la investigadora principal 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 IP.
- 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 AVPD con nº de registro: **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 la Responsable de 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)

English version

INFORMATION SHEET AND INFORMED CONSENT

The purpose of the present form is to provide the necessary information so that you can freely and voluntarily decide on the participation of your child in this research project. Please, read the following information carefully and do not hesitate to ask if you have any questions about it.

CONTACT:

Main researcher: María del Pilar García Mayo
Address: Paseo de la Universidad 5
Center: Facultad de Letras
Telephone: 945 013036
Email: mariapilar.garciamayo@ehu.eus

INFORMATION ABOUT THE PROJECT:

- **Project title**: *Interaction and written production. The potential of collaborative writing in the learning of English as a foreign language by primary school learners*
- **Funded by the Ministry of Economy and Competitiveness**
- **Project description**:
    - This study aims to explore the potential of collaborative writing in learning English as a foreign language and document the collaborative interaction between Primary Education students while they perform tasks that require the development of a written product.

DESCRIPTION OF THE PROCEDURE

- **Type of procedure**: participants will complete an oral task and a written production task
    - Both tasks will be collaborative tasks completed in pairs.
    - In the oral task, the participants will describe and narrate in pairs a story reflected in some vignettes where they will have to agree on the descriptions of each image.
    - In the written production task, the participants will complete a task similar to the previous one but this time, in addition to agreeing on the story, they must decide on a written production based on the narration.
    - Anonymous personal data: personal data will be treated completely anonymously, as well as the results of all tests.

- **Number of interventions**: data collection will be carried out in four sessions of approximately 15 minutes each in January and June at Colegio San Viator.
- **Description of the procedure**: In each session the participant will complete a collaborative task together with a partner with a time limit of approximately 15 minutes. The tasks are aimed at obtaining information about the benefits of collaborative tasks and receiving feedback in the written production of Primary Education students.

- **Description of risks:** there is no risk for the student.

RIGHTS OF THE PARTICIPANT:

- Participation in this study is voluntary and you can stop participating at any time, without causing any harm, communicating the intention to abandon the main researcher by email.
- If you grant permission for your child to collaborate in this project, once it is finished, you will have at your disposal all the information related to the results obtained in it, respecting the confidentiality of the participants. You can obtain the data by contacting the main researcher.
- Tests can also include collecting information through recordings:
    - I consent to the recording
    - I DO NOT consent to the recording
- The personal data that you have provided us for this research project will be treated with absolute confidentiality in accordance with the Data Protection Law. The personal data collected object of this investigation will be included in a file registered in the AVPD with registration number: INM 0089 under the name Collaborative writing and young learners. The person responsible for the file or treatment will be the UPV / EHU.You can consult at any time the data you have provided us or request that we rectify or cancel your data or simply not use it for any specific purpose of this investigation. The way to do this is by contacting the UPV / EHU LOPD Security Manager, Rectorate, Barrio Sarriena s / n, 48940 Leioa-Bizkaia. For more information on Data Protection, we recommend consulting our website www.ehu.es/babestu on the Internet.

IDENTIFICATION OF THE PERSON GIVING THE CONSENT

I (name and surname) …… with ID card…….. as father / mother / legal representative of (name and surname of the student) ……

DECLARE

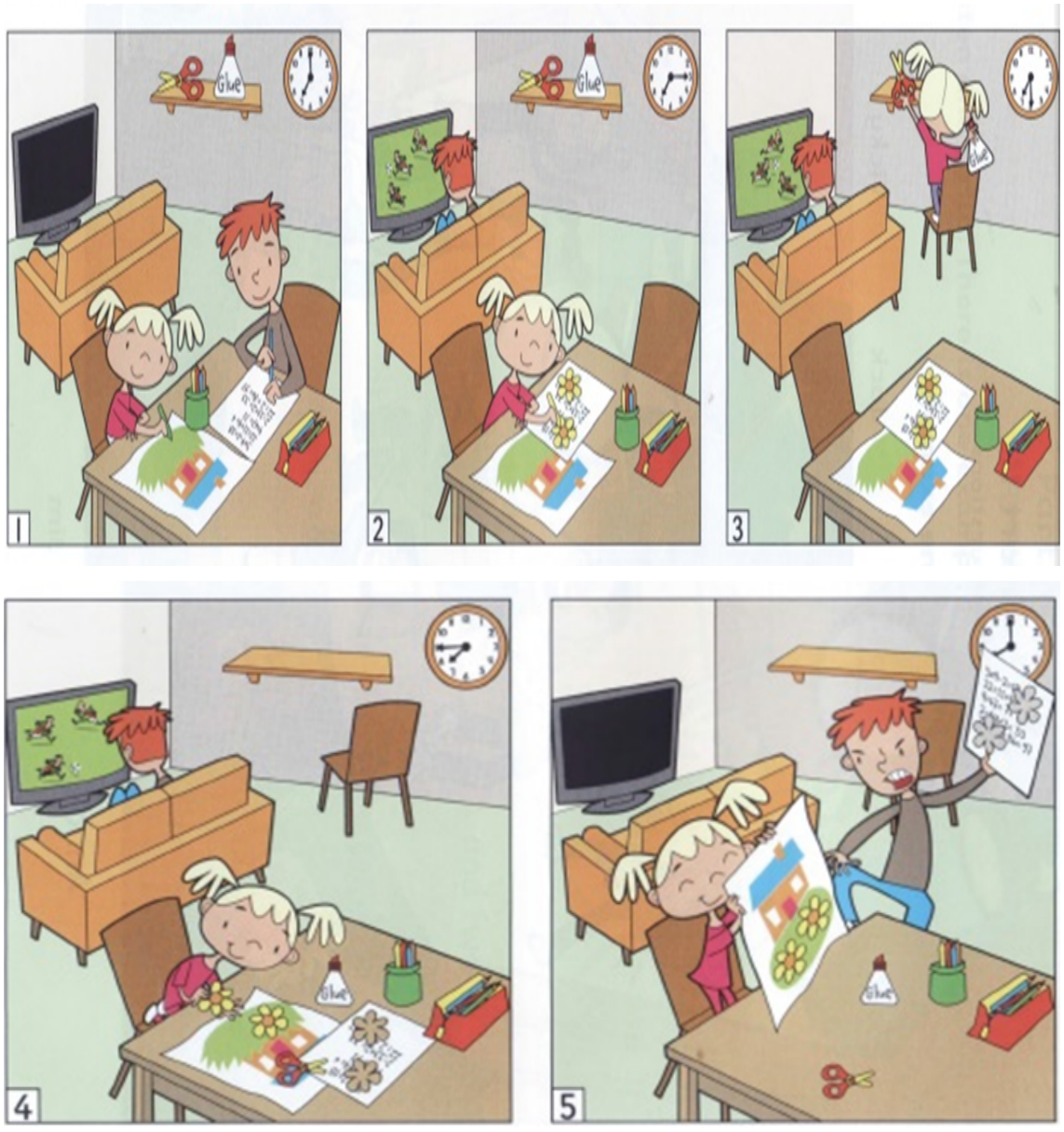
that I understand that this consent may be revoked by me at any time and **I GIVE MY CONSENT** to participate in this study.

(Date) (Signature of father/ mother / legal representative)

Appendix 3: Research design

	WEEK 1						WEEK 2			WEEK 3		
	Session 1			Session 2			Session 3			Session 4		
	CG (9 pairs)	TG (11 pairs)	LTG (10 pairs)	CG	TG	LTG	CG	TG	LTG	CG	TG	LTG
January CYCLE 1	Individual motivation questionnaire (5′) Stage 1 (15’): Production of a text based on picture prompt	Individual motivation questionnaire (5′) Stage 1 (15’): Production of a text based on picture prompt	Individual motivation questionnaire (5′) Stage 1 (15’): Production of a text based on picture prompt	Stage 2 (15’): Self-correction	Stage 2 (15’): Comparison with a model	Stage 2 (15’): Comparison with a model	Stage 3 (15’): Rewriting of original text Motivation questionnaire (5’)	Stage 3 (15’): Rewriting of original text Motivation questionnaire (5’)	Stage 3 (15’): Rewriting of original text Motivation questionnaire (5’)	Stage 4 (15’): Production of a new text based on a different picture prompt	Stage 4 (15’): Production of a new text based on a different picture prompt	Stage 4 (15’): Production of a new text based on a different picture prompt
January-June	The LTG (comparison at stage 2) and the CG (self-correction at stage 2) work with 4 pictures (and models) administered by their corresponding teachers.											
June CYCLE 2	Stage 1 (15’): Production of a text based on picture prompt	Stage 1 (15’): Production of a text based on picture prompt	Stage 1 (15’): Production of a text based on picture prompt	Stage 2 (15’): Self-correction	Stage 2 (15’): Comparison with a model	Stage 2 (15’): Comparison with a model	Stage 3 (15’): Rewriting of original text Motivation questionnaire (5’)	Stage 3 (15’): Rewriting of original text Motivation questionnaire (5’)	Stage 3 (15’): Rewriting of original text Motivation questionnaire (5’)	Stage 4 (15’): Production of a new text based on a different picture prompt Focus group (15’)	Stage 4 (15’): Production of a new text based on a different picture prompt Focus group (15’)	Stage 4 (15’): Production of a new text based on a different picture prompt Focus group (15’)

Appendix 4: The picture prompt used in the training session



(Taken from Cambridge English (2014). *Young Learners English Tests*)

Appendix 5: The model text used in the training session

The masterpiece

It's Monday evening and Tom is doing his homework while his sister Katie does some drawing. They are in the sitting room of their house. After a while, Tom gets bored and decides to watch football on TV. Katie has drawn a lovely house with a garden, but she wants to put some flowers on the grass. She sees Tom's homework and starts to draw flowers on it!

Soon Katie's flowers are finished, so she gets some scissors and glue, cuts them out and sticks them onto her picture. Proud of her work, Katie shows her picture to her brother. When Tom realises where the flowers have come from and sees his ruined homework, he is furious! But Katie is very happy with her picture.

(Taken from Cambridge English (2014). *Young Learners English Tests*)

Appendix 6: The pairs' original drafts used in the training session

Stage 1

Names:_____

Instructions:

1) Look at the set of pictures and write the story together!

2) You have 15 minutes

3) Speak in English!

One day two brothers they are doing homework at home.

When the young brother ginish his homework he go to see the tv, then the sister take the homework of his brother and paint two flowers

When the sister finish painting the flowers she go to take sisorsh and glue.

The sister take the sisodsh and cut the flowers and them put in the paintings.

When the brother finish watching the TV go to take his homework and see this homework broken.

Stage 1

Names:_____

Instructions:

1) Look at the set of pictures and write the story together!

2) You have 15 minutes

3) Speak in English!

In the first Picture, there are a boy making their homework and a girl painting in the living room

In the second picture, the boy is watching TV and the girl is painting in the boys homework.

In the third picture, the girl is taking the scissors and the glue.

In the fourth picture, the girl cut the pictures of the boys homework and is sticking the pictures in shes picture whit the glue.

In the last Picture, the girl has finish the picture and de boy is angry because the cut his homework.

Stage 1

Names: _____

Instructions:

1) Look at the set of pictures and write the story together!

2) You have 15 minutes

3) Speak in English!

At 7.00 it's Juana and peter. Peter is doing her homeworks in the table, Juana is ~~painting~~ a big house.

At 7.15 , Juana is doing flowers in the homeworks of peter, Peter is waching the tV.

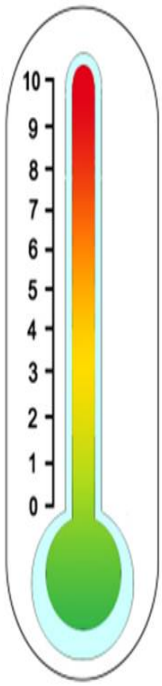
At 7.30 ,Juana is taking the ~~s~~ scissors and the glue.

At 7:45, Juana put the flowers and glue in

Appendix 7: The motivation thermometer

The Motivation Thermometer

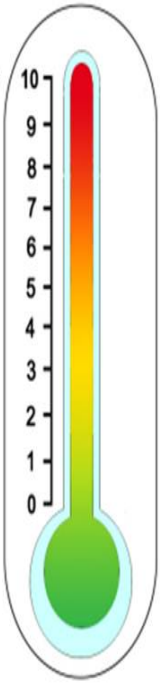
Name, school, grade: _____

¿Cómo te sientes **antes** de hacer la tarea?	¿Por qué has elegido este nivel?
	**Porque creo que la tarea va a ser fácil**. **Porque quiero trabajar con mi compañero/a**. **Porque tengo ganas de hacer la tarea**. **Porque quiero hacer una actividad en inglés.** **Porque creo que me voy a divertir haciendo la tarea.** **Porque creo que la tarea va a ser difícil.** **Porque no quiero trabajar con mi compañero/a.** **No tengo ganas de hacer la tarea.** **Porque no quiero hacer una actividad en inglés.** **Porque creo que me voy a aburrir haciendo la tarea.**

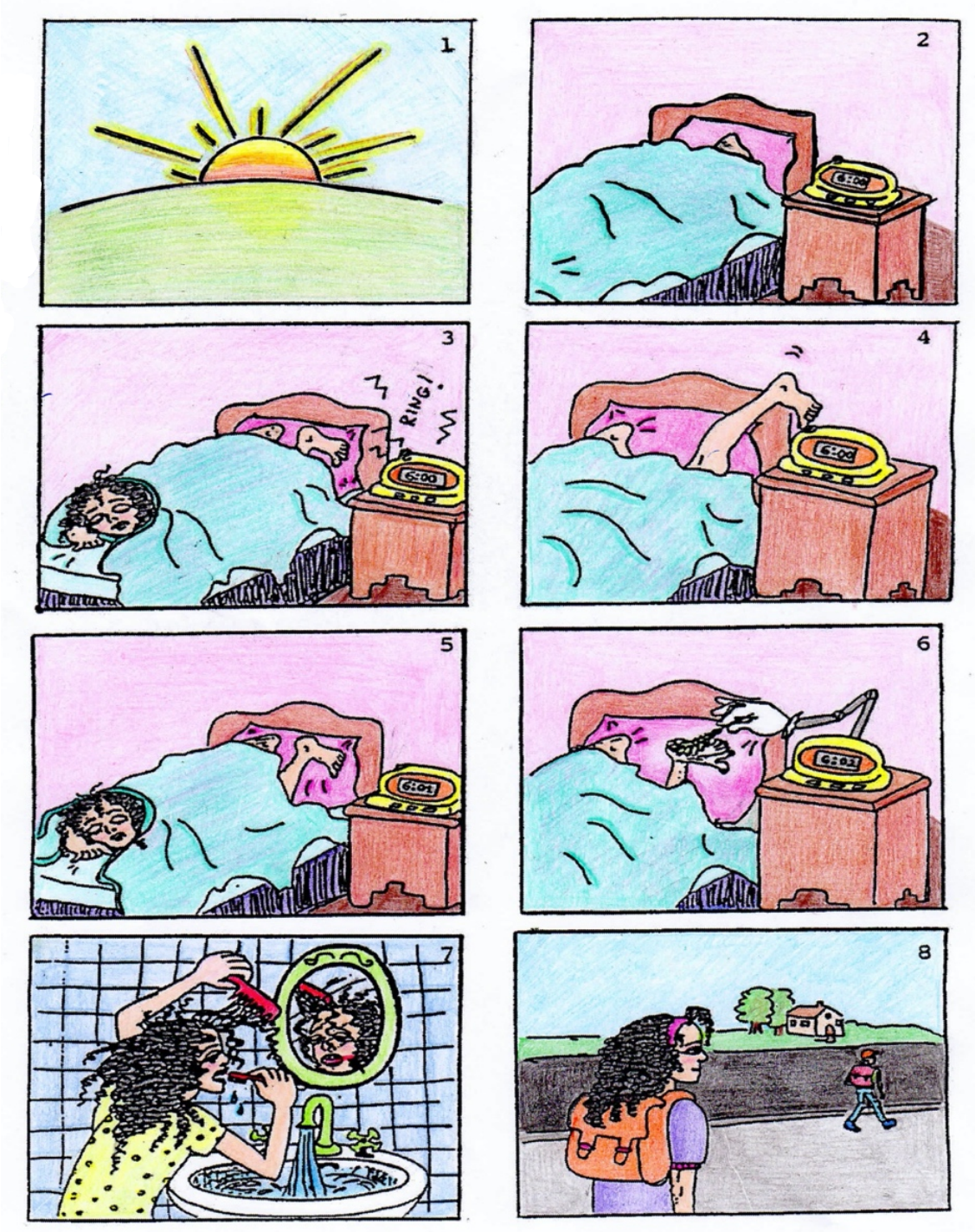
(Adapted from Al Khalil, 2016)

Note: prompts created by the researchers

English version

*¿How do you feel **before** doing the task?* *Please, indicate on this thermometer*	Why have you chosen this level? Please, indicate
	**Because I think the task is going to be easy.** **Because I want to work with my peer.** **Because I want to do the task.** **Because I want to do an activity in English.** **Because I think I´m going to have fun doing the task.** **Because I think the task is going to be difficult.** **Because I don't want to work with my peer.** **Because I don't want to do the task.** **Because I don't want to do the activity in English.** **Because I think I'm going to get bored doing the task.**

Appendix 8: 'Martine's alarm-clock' picture story



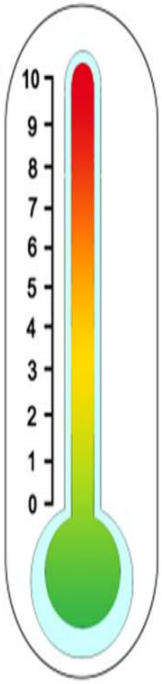
(Taken from Lapkin, Swain, & Smith, 2002)

Martine's alarm-clock

It's six a.m. and the sun is rising. Martine is sound asleep in her bed. She's having sweet dreams, her head at the foot of the bed and her feet on the pillow. When the alarm clock rings, Martine doesn't want to get up. She sticks her foot out, and with her big toe, she shuts off the alarm. She falls asleep again immediately. But she has the kind of alarm clock you need to prevent being late. At 6:02, a mechanical hand holding a small feather comes out of the alarm clock. It tickles her foot. To good effect! Finally, Martine gets up. She brushes her teeth, combs her hair and gets dressed to go to school. Another great start to the day!

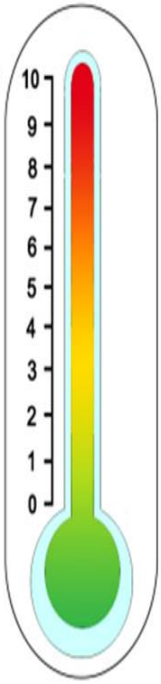
(Taken from Lapkin, Swain, & Smith, 2002)

Appendix 10: The post-task motivation thermometer

¿Cómo te sientes **después** de hacer la tarea?	¿Por qué has elegido este nivel?
	**Porque la tarea ha sido fácil.** **Porque he trabajado a gusto con mi compañero/a.** **Porque me ha gustado la tarea.** **Porque me ha gustado hacer una actividad en inglés.** **Porque me he divertido haciendo la tarea.** **Porque la tarea ha sido difícil.** **Porque no me ha gustado trabajar con mi compañero/a.** **Porque no me ha gustado la tarea.** **Porque no me gusta hacer actividades en inglés.** **Porque me he aburrido.**

(Taken from Al Khali, 2016)

Note: prompts created by the researchers

*¿How do you feel **after** doing the task?* *Please, indicate on this thermometer*	Why have you chosen this level? Please, indicate
	**Because the task was easy.** **Because I enjoyed working with my peer.** **Because I liked the task.** **Because I enjoyed doing an activity in English.** **Because I enjoyed doing the task.** **Because the task was difficult.** **Because I didn't enjoy working with my peer.** **Because I didn't like the task.** **Because I don't like doing activities in English.** **Because I got bored.**

Appendix 11: 'One more day to school' picture story

(©Israel Azpilicueta)

Appendix 12: 'The cat has six lives' picture story

(Taken from Cánovas Guirao, 2017)

THE CAT HAS SIX LIVES

At the weekend, three children with their dog go to the countryside to visit their grandmother and her cat.

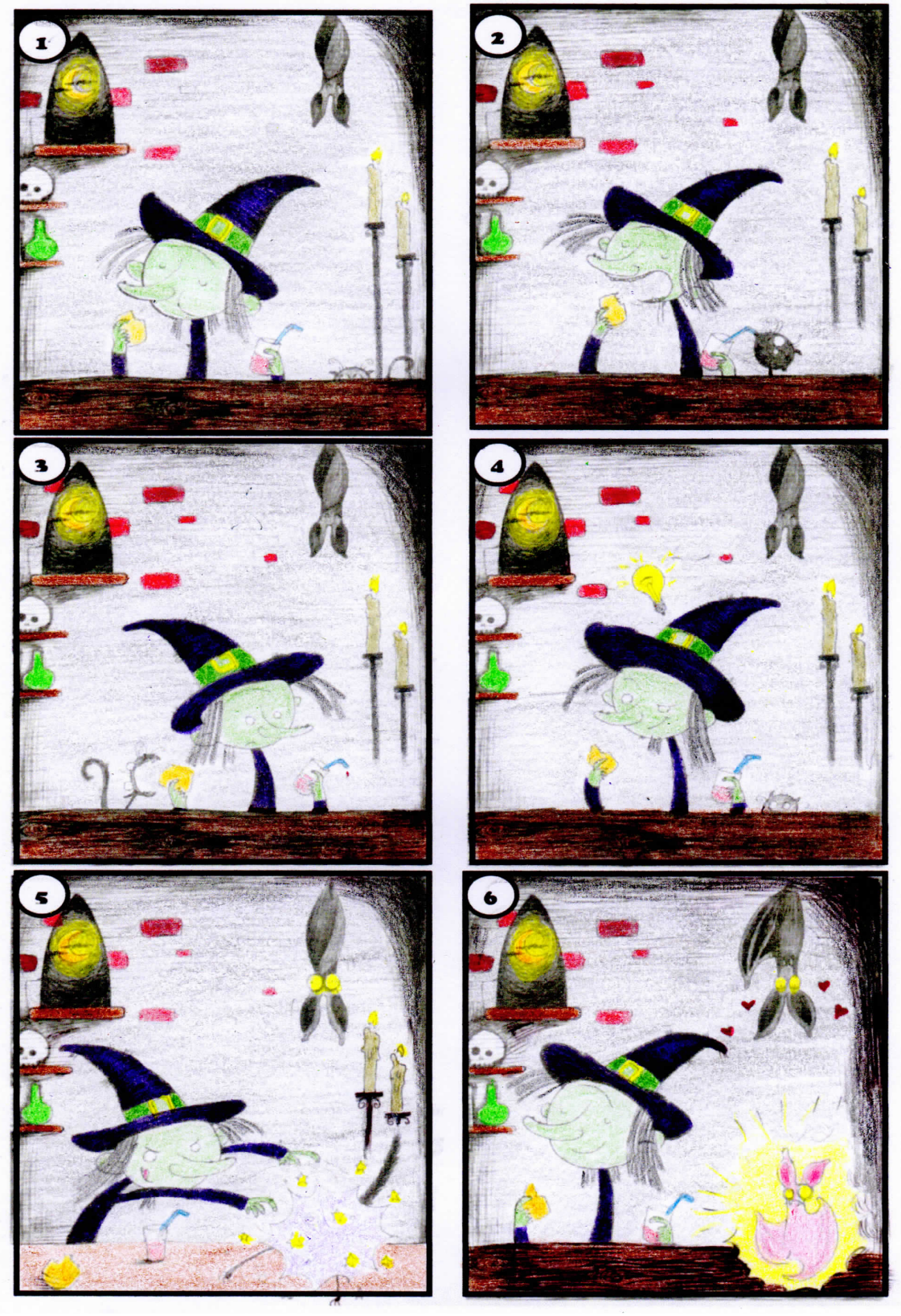
Suddenly, when the dog sees the cat, the dog wants to catch it. The cat is afraid of the dog and climbs up a tree. Up in the tree the cat looks at the dog angrily, the dog barks at the cat and the grandmother and her grandchildren look surprised and worried.

Then, the children want to save the cat because their grandmother is very worried. First, one of the boys draws a ladder on a piece of paper and, instantly, a magic ladder appears in the garden.

The boy climbs up the ladder, saves the cat and granny is so happy for her cat that gives everybody some apples.

(Adapted from Cánovas Guirao, 2017)

Appendix 14: 'The witch' picture story



(Taken from Cánovas Guirao, 2017)

THE WITCH

After conducting experiments, a witch is having dinner with her bat and her cat in her laboratory full of potions. She prepares a sandwich and a glass of orange juice. Then, her cat sees the food.

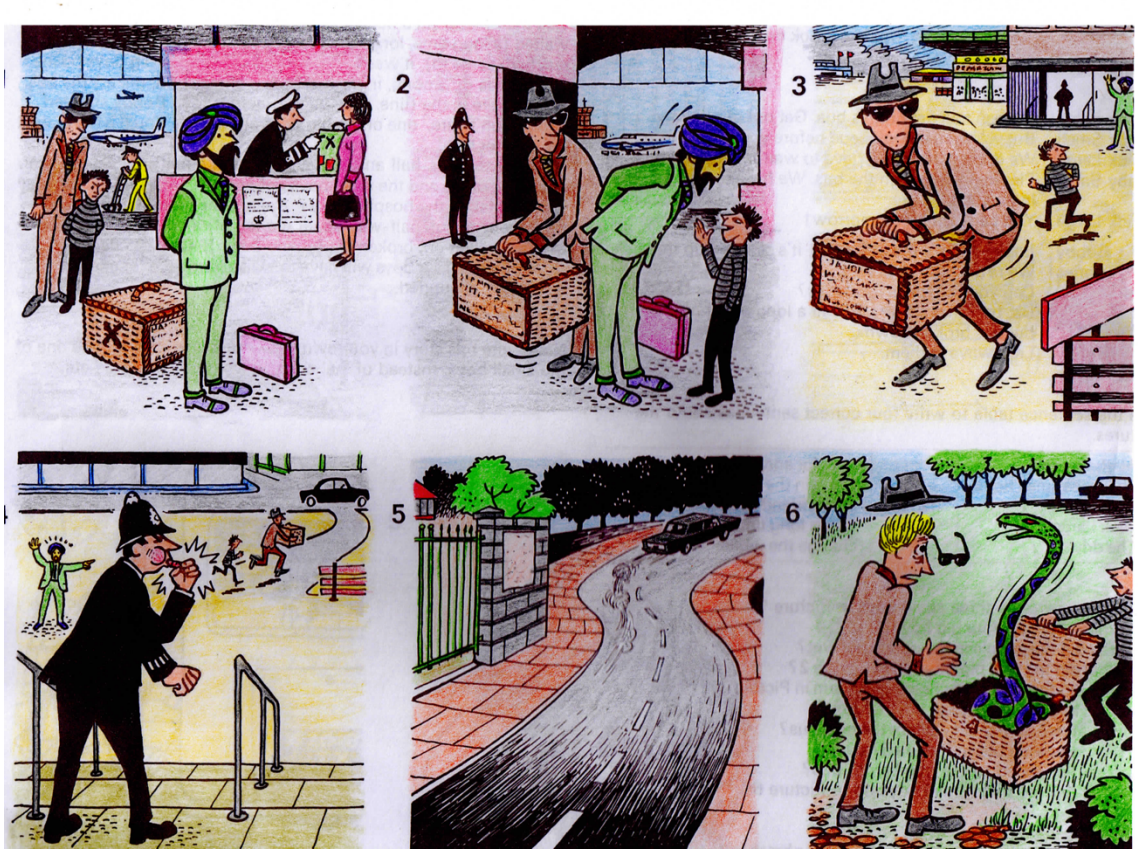
Then, while the witch is eating her sandwich, the cat drinks her orange juice. After that, the witch sees that some of her orange juice is missing. And when she is looking at her glass, the cat eats her sandwich, too.

Afterwards, the witch sees the cat and has an idea. The witch casts a spell. There is a bright flash of light and a loud sound!

Finally, there is a loud noise and a bright flash of light. The witch has turned the cat into a white bat! Now she can have her snack in peace. And the black bat wakes up and falls in love with the white bat.

(Adapted from Cánovas Guirao, 2017)

Appendix 16: 'Ssssurprise!' picture story



(Taken from the IRIS digital repository)

Ssssurprise!

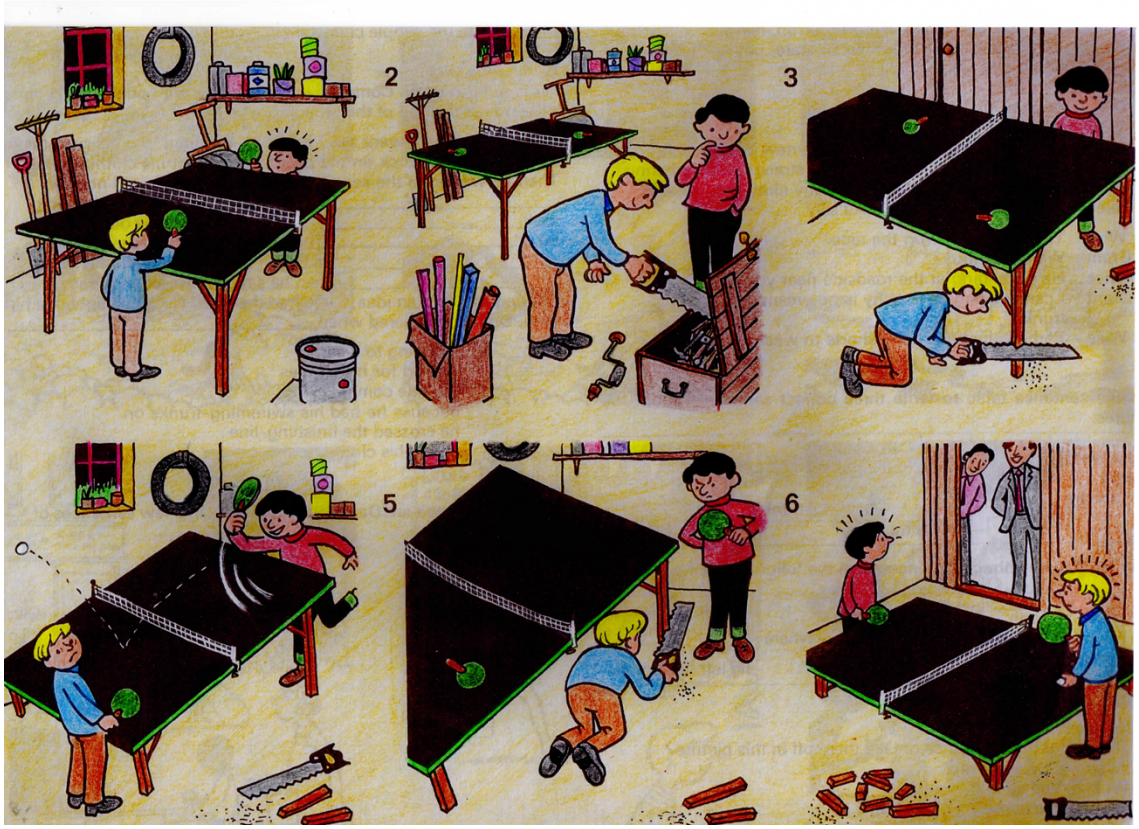
In the middle of a busy airport, the thief saw his next victim. A businessman dressed in a green suit was standing with a big basket and a small suitcase behind him.

The thief used his friend to distract the man. The thief took his basket and ran away with it, while the businessman cried for help.

The two criminals ran to a car while a policeman tried to stop them using his whistle, but the car drove off along the road as fast as it could.

The driver dropped the thieves off on the side of the road, in the middle of a wood. They got a big surprise when they opened the basket. There was no money, but out danced a big snake!

Appendix 18: 'The table that got smaller' picture story



(Taken from the IRIS digital repository)

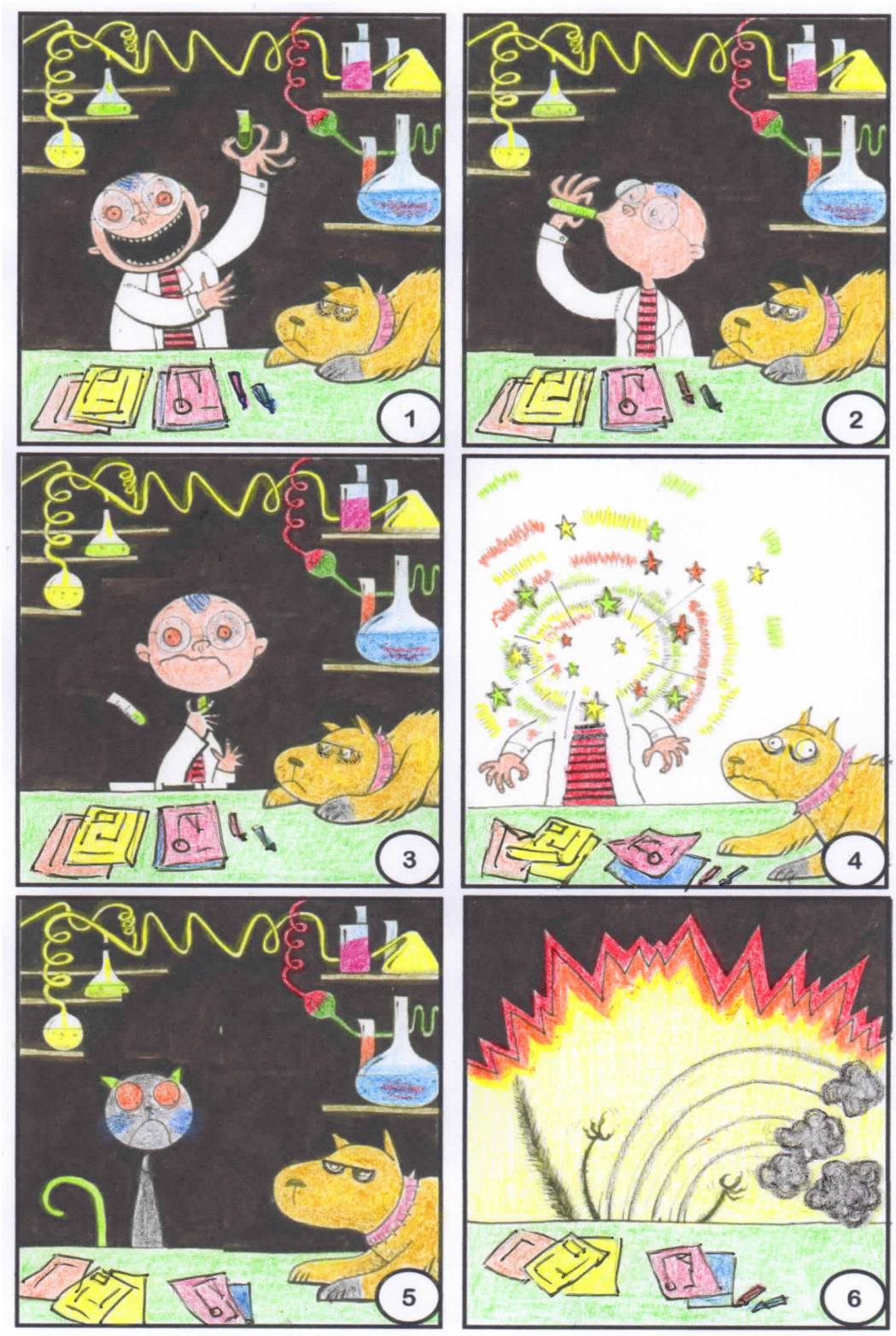
THE TABLE THAT GOT SMALLER

Billy and Mark were happy to try out the new ping pong table. But it was too high for them. Suddenly, Billy had an idea! He took a saw from his dad's tool box. He began sawing the legs on his side of the table.

Soon Billy realized that, as his side was so short, Mark would have an advantage while playing. Billy was not happy about that. He took his dad's saw and cut the legs on Mark's side of the table. This made Mark really angry. Billy cut the legs again and again. The table was too short now!

When Billy's and Mark's dads appeared, the boys were worried that they got angry, but their dads just laughed at what their sons had done!

Appendix 20: 'The scientist' picture story



(Taken from Cánovas Guirao, 2017)

The scientist

One day, a scientist is in his laboratory conducting experiments. It´s late and his dog is sleeping on the table next to him. After hours of work, he finishes his new potion. He is very excited and decides to test it immediately. He drinks the whole potion. Suddenly, he feels very strange. There is a loud noise and a bright flash of light. His dog wakes up terrified. Then, the scientist turns into a cat! Finally, the scientist's dog looks angrily at the cat and attacks him.

(Taken from Cánovas Guirao, 2017)

Appendix 22: 'The bat scientist' picture story

(©Israel Azpilicueta)

Appendix 23: Attitudes toward the task questionnaire (CG)

Cuestionario

Contesta a las siguientes preguntas eligiendo la respuesta correcta o dando tu opinión. Recuerda que el cuestionario es **anónimo**.

1. Da tu opinión sobre la siguiente actividad: **Escribir un texto a partir de unos dibujos.**

   a. ¿Sueles hacer este tipo de actividades en tus clases de inglés del colegio?

   Sí ◯ A veces ◯ No ◯

   b. ¿Has disfrutado con la actividad?

En absoluto	No mucho	Así así	Bastante	Mucho
◯	◯	◯	◯	◯

   ¿Por qué? _____

   c. ¿Crees que puede ser útil para mejorar tu inglés?

En absoluto	No mucho	Así así	Bastante	Mucho
◯	◯	◯	◯	◯

   ¿Por qué? _____

   d. ¿Te gustaría que tu profesor/a continuara realizando este tipo de actividad?

   Sí ◯ No ◯

   ¿Por qué? _____

2. Da tu opinión sobre la siguiente actividad: **Autocorrección.**

a. ¿Sueles hacer este tipo de actividades en tus clases de inglés del colegio?

Sí ◯ A veces ◯ No ◯

b. ¿Has disfrutado con la actividad?

En absoluto	No mucho	Así así	Bastante	Mucho
◯	◯	◯	◯	◯

¿Por qué? _____

c. ¿Crees que puede ser útil para mejorar tu inglés?

En absoluto	No mucho	Así así	Bastante	Mucho
◯	◯	◯	◯	◯

¿Por qué? _____

d. ¿Te gustaría que tu profesor/a continuara realizando este tipo de actividad?

Sí ◯ No ◯

¿Por qué? _____

Escribe aquí cualquier otra cosa que te gustaría añadir: sugerencias, comentarios, etc.:_____
_____
_____
_____
_____

Gracias por tu colaboración

Cuestionario

Contesta a las siguientes preguntas eligiendo la respuesta correcta o dando tu opinión. Recuerda que el cuestionario es **anónimo**.

1. Da tu opinión sobre la siguiente actividad: **Escribir un texto a partir de unos dibujos.**

   a. ¿Sueles hacer este tipo de actividades en tus clases de inglés del colegio?

   Sí ◯ A veces ◯ No ◯

   b. ¿Has disfrutado con la actividad?

En absoluto	No mucho	Así así	Bastante	Mucho
◯	◯	◯	◯	◯

   ¿Por qué? _____

   c. ¿Crees que puede ser útil para mejorar tu inglés?

En absoluto	No mucho	Así así	Bastante	Mucho
◯	◯	◯	◯	◯

   ¿Por qué? _____

   d. ¿Te gustaría que tu profesor/a continuara realizando este tipo de actividad?

   Sí ◯ No ◯

   ¿Por qué? _____

2. Da tu opinión sobre la siguiente actividad: **Comparar tu texto con un modelo.**

a. ¿Sueles hacer este tipo de actividades en tus clases de inglés del colegio?

Sí ◯ A veces ◯ No ◯

b. ¿Has disfrutado con la actividad?

En absoluto	No mucho	Así así	Bastante	Mucho
◯	◯	◯	◯	◯

¿Por qué? _____

c. ¿Crees que puede ser útil para mejorar tu inglés?

En absoluto	No mucho	Así así	Bastante	Mucho
◯	◯	◯	◯	◯

¿Por qué? _____

d. ¿Te gustaría que tu profesor/a continuara realizando este tipo de actividad?

Sí ◯ No ◯

¿Por qué? _____

Escribe aquí cualquier otra cosa que te gustaría añadir: sugerencias, comentarios, etc.:_____
_____
_____
_____
_____

Gracias por tu colaboración

(Taken from Loidi Labandibar, 2016)

Questionnaire

Please answer the questions ***by choosing the right answer*** *or* ***by explaining your thoughts as fully as possible****, where necessary. Remember that the questionnaire is* ***anonymous****.*

1. Give your opinion on the following activity: **Writing a text in response to a picture**

a. Do you usually do activities of this type in your English classes at school?

Yes ◯ Sometimes ◯ No ◯

b. Did you enjoy the activity?

Not at all	Not so much	So-so	Quite a lot	Very much
◯	◯	◯	◯	◯

Why? _____

c. Do you think it might be useful to improve your English?

Not at all	Not so much	So-so	Quite a lot	Very much
◯	◯	◯	◯	◯

Why? _____

d. Would you like your teacher to continue doing this type of activity?

Yes ◯ No ◯

Why? _____

2. Give your opinion on the following activity: **Comparing your text with a model/self-correcting your text.**

a. Do you usually do activities of this type in your English classes at school?

Yes ◯ Sometimes ◯ No ◯

b. Did you enjoy the activity?

Not at all	Not so much	So-so	Quite a lot	Very much
◯	◯	◯	◯	◯

Why? _____

c. Do you think it might be useful to improve your English?

Not at all	Not so much	So-so	Quite a lot	Very much
◯	◯	◯	◯	◯

Why? _____

d. Would you like your teacher to continue doing this type of activity?

Yes ◯ No ◯

Why? _____

Write here anything else you would like to add -suggestions, comments, etc.:_____
_____
_____
_____
_____

Thank you for your collaboration

Appendix 24: Interview questions

FOCUS GROUP INTERVIEW

1. ¿Habíais hecho esta actividad alguna vez? ¿Os gustaría hacer esto en clase?
2. ¿Cómo creíais que iba a ser esta actividad?
3. ¿Os parece útil? ¿Para qué?
4. ¿Cómo os gustaría que os corrigieran?
5. ¿Os ha gustado hacer esta actividad? ¿Qué os ha gustado? ¿Qué no os ha gustado? ¿Cambiaríais algo?
6. ¿Habéis mejorado vuestro inglés? ¿En qué habéis mejorado?
7. ¿Qué os ha parecido comparar vuestro texto con un modelo/autocorregiros? ¿Creéis que es útil para mejorar vuestra escritura?
8. ¿Habéis mantenido vuestra motivación durante todo el proceso o la habéis perdido en algún momento?
9. ¿Os ha gustado trabajar con vuest@ compañer@? ¿Por qué?
10. ¿Hubierais preferido hacerlo solos? ¿Por qué?

English version

FOCUS GROUP INTERVIEW

1. Have you ever done this activity? Would you like to do it in class?

2. How did you think this activity was going to be like?

3. Do you find it useful? For what?

4. How would you like to be corrected?

5. Did you like doing this activity? What did you like? What did you not like? Would you change something?

6. Do you think your English is better now? In which ways?

7. What is your opinion on comparing your text with a model/self-correcting your text? Do you think it is useful to improve your writing?

8. Have you been motivated throughout the whole process, or have you lost your motivation at some point?

9. Did you like working with your partner? Why?

10. Would you have preferred to do it alone? Why?

Appendix 25: Classification of LREs in previous studies

CATEGORIES	REFERENCES
Content-related episodes (CREs)	Cánovas et al. (2015) Yang & Zhang (2010) Abe (2008)
Content	Hanaoka (2006a, 2006b, 2007) Hanaoka & Izumi (2012) Kang (2020) Luquin & García Mayo (2020)
Content and discourse	García Mayo & Loidi Labandibar (2017)
Ideas and expressions	Lázaro-Ibarrola (2021) Coyle & Roca de Larios (2014) Martínez Esteban & Roca de Larios (2010) Montealegre Ramón (2019)
Discourse (D-LRE): logic sequencing such as cohesion and coherence, achieving inter-sentential relationship, organization of paragraphs, cohesive devices, and stylistics	Cánovas et al. (2015) Qi & Lapkin (2001) Coyle et al. (2018) Yang & Zhang (2010) Abe (2008) Montealegre Ramón (2019)
Mechanics (spelling+punctuation+pronunciation)	Calzada & García Mayo (2020a) Luquin & García Mayo (2020)
Spelling	Coyle & Roca de Larios (2014) Martínez Esteban & Roca de Larios (2010) Lázaro-Ibarrola (2021)
Sentence (S-LRE): whole sentences/clauses	Cánovas et al. (2015) Coyle & Roca de Larios (2020)
Lexis (L-LRE): verb, adjective, adverb, preposition, noun, pronoun, noun phrase	Cánovas et al. (2015) Qi & Lapkin (2001) Coyle et al. (2018) Yang & Zhang (2010) Abe (2008) Montealegre Ramón (2019) Lázaro-Ibarrola (2021) Coyle & Roca de Larios (2014) Hanaoka (2006a, 2006b, 2007) Martínez Esteban & Roca de Larios (2010) García Mayo & Loidi Labandibar (2017) Hanaoka & Izumi (2012) Kang (2020) Luquin & García Mayo (2020) Coyle & Roca de Larios (2020)
Lexis (meaning+word choices+prepositions+spelling)	Basterrechea & Leeser (2019) Collins & White (2019)
Meaning-focused	Calzada & García Mayo (2020a)
Form (F-LRE): verb form, verb tense, sentence structure, spelling, subject-	Cánovas et al. (2015) Qi & Lapkin (2001)

verb agreement, plural, poss. marker, poss. pronoun, comparative, superlative, punctuation	Coyle et al. (2018) Yang & Zhang (2010) Abe (2008) Kang (2020) Montealegre Ramón (2019) Luquin & García Mayo (2020)
Grammar: morphology: tense, part of speech, plural or pronoun; + syntax: word order in questions, adjective/noun order, article choice	Basterrechea & Leeser (2019) Collins & White (2019) Lázaro-Ibarrola (2021) Coyle & Roca de Larios (2014) Hanaoka (2006a, 2006b, 2007) Martínez Esteban & Roca de Larios (2010) García Mayo & Loidi Labandibar (2017) Hanaoka & Izumi (2012)
Form-focused: F-target or F-others	Calzada & García Mayo (2020a)
Form: Verb form, verb tense and spelling	Coyle & Roca de Larios (2020)
Organization	Kang (2020)
Pronunciation	Collins & White (2019)
Other	Lázaro-Ibarrola (2021) García Mayo & Loidi Labandibar (2017) Martínez Esteban & Roca de Larios (2010) Abe (2008) Hanaoka (2006a) Hanaoka & Izumi (2012) Kang (2020) Luquin & García Mayo (2020)

Appendix 26: Rubric for marking the writing

		MARKS		
		3	2	1
TASK	ADEQUACY	All the points in the instructions are mentioned; all the parts of a story are included (beginning, body, ending); the length of the text is appropriate.	Just some points in the instructions are mentioned; most of the parts of a story are included; the text is too short (ideas are not fully developed).	Notable omission 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 organised (use of paragraphs). Cohesive devices linking sentences and paragraphs. No serious mistakes.	Ideas are organised. Some cohesive devices linking sentences and paragraphs. There may be some mistakes.	There is a lack of organisation or linking devices.
	GRAMMATICAL ACCURACY	Very few, irrelevant, or no grammar errors at all. Good command of grammar; use of Past tense.	Some acceptable grammar errors. Fair command of English grammar; use of Past tense (with some mistakes)	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.

Appendix 27: Itemization of PFNs at Stage 1, Cycle 1

	WRITING STAGE																													
	CREs	LREs																												
		MECHANICS			LEXIS								FORM													DISCOURSE			OTHER	
CYCLE 1		SPELLING	PUNCTUATION	PRONUNCIATION	VERB	ADJECTIVE	ADVERB	PREPOSITION	NOUN	PRONOUN	NOUN PHRASE	VERB PHRASE	VERB FORM (agreement)	VERB TENSE	SUBJECT	OBJECT	PLURAL	POSSESSIVE MARKER	POSSESSIVE PRONOUN	COMPARATIVE & SUPERLATIVE	PRONOUN	ADVERB	WORD ORDER	ARTICLE CHOICE	PREPOSITION	COHESION & COHERENCE	PARAGRAPHS	STYLISTICS		TOTAL PFN
GRADE 6A - CONTROL GROUP	12	22	18	2	15	0	0	0	23	0	0	8	13	11	1	0	2	1	4	0	6	2	1	2	3	0	1	5	2	154
%	7,79	14,286	11,688	1,30	9,7403	0	0	0	14,935	0	0	5,19	8,44	7,14	0,65	0	1,30	0,65	2,60	0	3,90	1,30	0,65	1,30	1,95	0	0,65	3,25	1,30	100
GRADE 6B - TREATMENT GROUP	27	22	14	1	24	4	5	0	23	2	0	5	19	5	1	0	3	2	12	0	3	0	0	0	4	2	3	4	1	186
%	14,52	11,828	7,53	0,54	12,903	2,15	2,69	0	12,366	1,08	0	2,69	10,22	2,69	0,54	0	1,61	1,08	6,45	0	1,61	0	0	0	2,15	1	1,61	2,15	0,54	100
GRADE 6C - DOUBLE TREATMENT GROUP	17	22	25	0	19	2	3	0	14	0	0	2	19	10	4	0	6	0	6	0	4	0	1	1	6	2	4	5	1	173
%	9,83	12,717	14,45	0	10,983	1,16	1,73	0	8,09	0	0	1,16	10,98	5,78	2,31	0	3,47	0	3,47	0	2,31	0	1	1	3,47	1	2,31	2,89	0,58	100

Appendix 28: Itemization of PFNs at Stage 1, Cycle 2

	WRITING STAGE																													
	CREs	LREs																												
		MECHANICS			LEXIS								FORM												DISCOURSE			OTHER	
CYCLE 2		SPELLING	PUNCTUATION	PRONUNCIATION	VERB	ADJECTIVE	ADVERB	PREPOSITION	NOUN	PRONOUN	NOUN PHRASE	VERB PHRASE	VERB FORM (agreement)	VERB TENSE	SUBJECT	OBJECT	PLURAL	POSSESSIVE MARKER	POSSESSIVE PRONOUN	COMPARATIVE & SUPERLATIVE	PRONOUN	ADVERB	WORD ORDER	ARTICLE CHOICE	PREPOSITION	COHESION & COHERENCE	PARAGRAPHS	STYLISTICS		TOTAL PFN
GRADE 6A - CONTROL GROUP	6	32	18	4	14	0	3	0	11	0	0	0	9	9	1	0	1	0	10	0	5	0	0	2	3	1	1	1	4	135
%	4,44	23,704	13,333	2,96	10,37	0	2,22	0	8,15	0	0	0	6,67	6,67	0,74	0	0,74	0	7,41	0	3,70	0	0	1,48	2,22	1	0,74	0,74	2,96	100
GRADE 6B - TREATMENT GROUP	16	19	12	4	22	1	3	0	10	0	1	0	4	14	1	0	0	0	5	0	4	1	1	2	4	2	3	6	2	137
%	11,68	13,869	8,76	2,92	16,058	0,73	2,19	0	7,30	0	0,7299	0	2,92	10,22	0,73	0	0	0	3,65	0	2,92	1	1	1	2,92	1	2,19	4,38	1,46	100
GRADE 6C - DOUBLE TREATMENT GROUP	3	15	5	1	15	3	0	0	3	0	0	1	2	11	2	0	0	0	3	0	4	0	1	1	2	1	5	8	2	88
%	3,41	17,045	5,68	1	17,045	3,41	0	0	3,41	0	0	1,14	2,27	12,50	2,27	0	0	0	3,41	0	4,55	0	1	1	2,27	1	5,68	9,09	2,27	100

Appendix 29: Itemization of FNs at Stage 2, Cycle 1:

	COMPARISON STAGE																													
	CREs	LREs																												
		MECHANICS			LEXIS								FORM												DISCOURSE			OTHER	
CYCLE 1		SPELLING	PUNCTUATION	PRONUNCIATION	VERB	ADJECTIVE	ADVERB	PREPOSITION	NOUN	PRONOUN	NOUN PHRASE	VERB PHRASE	VERB FORM (agreement)	VERB TENSE	SUBJECT	OBJECT	PLURAL	POSSESSIVE MARKER	POSSESSIVE PRONOUN	COMPARATIVE & SUPERLATIVE	PRONOUN	ADVERB	WORD ORDER	ARTICLE CHOICE	PREPOSITION	COHESION & COHERENCE	PARAGRAPHS	STYLISTICS		TOTAL PFN
GRADE 6A - CONTROL GROUP	2	11	1	0	1	0	0	0	3	0	0	2	6	5	1	1	1	1	2	0	0	0	1	1	3	4	0	5	0	51
%	3,92	21,569	1,96	0	1,96	0	0	0	5,88	0	0	3,92	11,76	9,80	1,96	2	1,96	1,96	3,92	0	0	0	1,96	1,96	5,88	8	0	9,80	0	100
GRADE 6B - TREATMENT GROUP	42	2	2	0	21	4	2	0	8	0	2	0	2	3	0	0	4	0	2	0	0	0	0	0	1	1	3	8	11	118
%	35,59	1,69	1,69	0	17,80	3,39	1,69	0	6,78	0	1,69	0	1,69	2,54	0	0	3,39	0	1,69	0	0	0	0	0	0,85	1	3	6,78	9	100
GRADE 6C - DOUBLE TREATMENT GROUP	37	2	3	0	18	3	0	0	16	0	0	1	1	2	0	0	4	0	6	0	0	0	0	0	2	0	3	5	5	108
%	34,26	1,85	2,78	0	16,67	2,78	0	0	14,81	0	0	1	0,93	1,85	0	0	3,70	0	5,56	0	0	0	0	0	1,85	0	3	4,63	5	100

Appendix 30: Itemization of FNs at Stage 2, Cycle 2:

	COMPARISON STAGE																													
	CREs	LREs																												
		MECHANICS			LEXIS								FORM												DISCOURSE			OTHER	
CYCLE 2		SPELLING	PUNCTUATION	PRONUNCIATION	VERB	ADJECTIVE	ADVERB	PREPOSITION	NOUN	PRONOUN	NOUN PHRASE	VERB PHRASE	VERB FORM (agreement)	VERB TENSE	SUBJECT	OBJECT	PLURAL	POSSESSIVE MARKER	POSSESSIVE PRONOUN	COMPARATIVE & SUPERLATIVE	PRONOUN	ADVERB	WORD ORDER	ARTICLE CHOICE	PREPOSITION	COHESION & COHERENCE	PARAGRAPHS	STYLISTICS		TOTAL PFN
GRADE 6A - CONTROL GROUP	4	11	9	0	4	0	0	0	1	0	0	0	3	2	2	0	1	0	1	0	0	0	0	2	3	0	0	6	3	52
%	7,69	21,154	17,31	0	7,69	0	0	0	1,92	0	0	0	5,77	3,85	3,85	0	1,92	0	1,92	0	0	0	0	3,85	5,77	0	0	11,54	6	100
GRADE 6B - TREATMENT GROUP	29	18	6	1	13	2	4	0	6	0	2	0	1	4	0	0	0	0	5	0	0	1	0	0	3	1	1	4	10	111
%	26,13	16,22	5,41	1	11,71	1,80	3,60	0	5,41	0	1,80	0	0,90	3,60	0	0	0	0	4,50	0	0	1	0	0	2,70	1	1	3,60	9	100
GRADE 6C - DOUBLE TREATMENT GROUP	18	10	3	0	7	2	0	0	4	0	0	0	8	1	0	0	0	0	5	0	0	0	0	1	1	1	0	2	8	71
%	25,35	14,08	4,23	0	9,86	2,82	0	0	5,63	0	0	0	11,27	1,41	0	0	0	0	7,04	0	0	0	0	1	1,41	1	0	2,82	11	100

Appendix 31: Raw frequencies of the outcome of the FNIs and FUIs from Stage 3, Cycle 1 per subcategories

		Correct			Incorrect		
		CG	TG	LTG	CG	TG	LTG
CRE	Content	**1**	**17**	**14**	**0**	**1**	**5**
Mechanics	Spelling	2	5	7	2	2	1
	Punctuation	0	0	1	0	0	0
	Total	**2**	**5**	**8**	**2**	**2**	**1**
Lexis	Verb	0	6	6	1	2	4
	Adjective	0	3	0	0	4	1
	Adverb	1	0	0	0	0	1
	Preposition	0	0	0	0	0	0
	Noun	1	9	11	1	0	3
	Pronoun	0	0	0	0	0	0
	NP	0	0	1	0	0	0
	VP	0	2	0	0	0	0
	Total	**2**	**20**	**18**	**2**	**6**	**9**
Form	Verb form	2	2	0	0	0	0
	Verb tense	4	2	2	1	0	0
	Subject	1	0	0	0	0	0
	Object	1	0	0	0	0	0
	Plural	0	3	2	0	0	1
	Poss. marker	0	0	0	0	0	0
	Poss. pronoun	0	2	6	1	1	0
	Comparative and superlative	0	0	0	0	0	0
	Pronoun	0	1	0	0	0	0
	Adverb	0	0	0	0	0	0
	Word order	0	0	0	0	0	0
	Article	0	0	0	0	0	0
	Preposition	0	1	1	0	0	0
	Total	**8**	**11**	**11**	**2**	**1**	**1**
Discourse	Coherence and cohesion	0	0	0	0	0	0
	Paragraphs	0	1	1	0	0	0
	Sylistics	0	2	4	0	0	4
	Total	**0**	**3**	**5**	**0**	**0**	**4**
Other	Other	**0**	**1**	**0**	**0**	**0**	**0**
Total		**13**	**57**	**56**	**6**	**10**	**20**

*Note that the pronunciation subcategory has been removed

Appendix 32: Raw frequencies of the outcome of the FNIs and FUIs from Stage 3, Cycle 2 per subcategories

		Correct			Incorrect		
		CG	TG	LTG	CG	TG	LTG
CRE	Content	**2**	**21**	**26**	**0**	**1**	**0**
Mechanics	Spelling	1	11	12	2	4	2
	Puntuation	2	2	2	0	0	0
	Total	**3**	**13**	**14**	**2**	**4**	**2**
Lexis	Verb	2	6	12	0	3	5
	Adjective	0	1	3	0	1	0
	Adverb	0	2	7	0	0	2
	Preposition	0	0	0	0	0	0
	Noun	0	5	4	0	1	0
	Pronoun	0	0	0	0	0	0
	NP	0	1	0	0	0	0
	VP	0	1	0	0	0	0
	Total	**2**	**16**	**26**	**0**	**5**	**7**
Form	Verb form	1	0	2	0	0	0
	Verb tense	0	3	3	0	0	0
	Subject	0	0	0	0	0	0
	Object	0	0	0	0	0	0
	Plural	0	0	0	0	0	0
	Poss. marker	0	0	0	0	0	0
	Poss. pronoun	0	0	7	0	1	0
	Comparative and superlative	0	0	0	0	0	0
	Pronoun	0	0	0	0	0	0
	Adverb	0	0	0	0	0	0
	Word order	0	0	0	0	0	0
	Article	0	0	0	0	0	0
	Preposition	0	1	2	1	1	0
	Total	**1**	**4**	**14**	**1**	**2**	**0**
Discourse	Coherence and cohesion	0	0	0	0	0	0
	Paragraphs	0	0	3	0	0	0
	Sylistics	1	7	4	0	0	0
	Total	**1**	**7**	**7**	**0**	**0**	**0**
Other	Other	**0**	**0**	**1**	**0**	**0**	**0**
Total		**9**	**61**	**88**	**3**	**12**	**9**

Appendix 33: Raw frequencies of the outcome of the FNIs and FUIs from Stage 4, Cycle 1 per subcategories

		Correct			Incorrect		
		CG	TG	LTG	CG	TG	LTG
CRE	Content	**0**	**1**	**3**	**0**	**0**	**1**
Mechanics	Spelling	0	2	4	1	0	1
	Puntuation	0	0	0	0	0	0
	Total	**0**	**2**	**4**	**1**	**0**	**1**
Lexis	Verb	0	3	3	0	4	3
	Adjective	0	0	0	0	0	0
	Adverb	0	0	0	0	0	0
	Preposition	0	0	0	0	0	0
	Noun	0	0	0	0	0	0
	Pronoun	0	0	0	0	0	0
	NP	0	0	0	0	0	0
	VP	0	0	0	0	0	0
	Total	**0**	**3**	**3**	**0**	**4**	**3**
Form	Verb form	0	0	0	0	0	0
	Verb tense	0	1	0	0	0	0
	Subject	0	0	0	0	0	0
	Object	0	0	0	0	0	0
	Plural	0	0	1	0	0	0
	Poss. marker	0	0	0	0	0	0
	Poss. pronoun	0	1	3	2	0	1
	Comparative and superlative	0	0	0	0	0	0
	Pronoun	0	0	0	0	0	0
	Adverb	0	0	0	0	0	0
	Word order	0	0	0	0	0	0
	Article	0	0	0	0	0	0
	Preposition	0	0	1	0	0	0
	Total	**0**	**2**	**5**	**2**	**0**	**1**
Discourse	Coherence and cohesion	0	0	0	0	0	0
	Paragraphs	0	0	0	0	0	1
	Sylistics	0	3	1	0	0	2
	Total	**0**	**3**	**1**	**0**	**0**	**3**
Other	Other	**0**	**0**	**0**	**0**	**0**	**0**
Total		**0**	**11**	**16**	**3**	**4**	**9**

Appendix 34: Raw frequencies of the outcome of the FNIs and FUIs from Stage 4, Cycle 2 per subcategories

		Correct			Incorrect		
		CG	TG	LTG	CG	TG	LTG
CRE	Content	**1**	**5**	**5**	**0**	**1**	**0**
Mechanics	Spelling	1	4	5	1	3	2
	Puntuation	1	1	1	0	0	0
	Total	**2**	**5**	**6**	**1**	**3**	**2**
Lexis	Verb	2	7	9	1	0	2
	Adjective	0	0	1	0	0	0
	Adverb	0	1	6	0	0	2
	Preposition	0	0	0	0	0	0
	Noun	1	3	4	0	2	0
	Pronoun	0	0	0	0	0	0
	NP	0	0	0	0	0	0
	VP	0	0	1	0	0	0
	Total	**3**	**11**	**21**	**1**	**2**	**4**
Form	Verb form	0	0	1	0	0	0
	Verb tense	0	0	0	0	0	0
	Subject	0	0	0	1	0	0
	Object	0	0	0	0	0	0
	Plural	0	0	0	0	0	0
	Poss. marker	0	0	0	0	0	0
	Poss. pronoun	0	1	5	0	1	0
	Comparative and superlative	0	0	0	0	0	0
	Pronoun	0	0	0	0	0	0
	Adverb	0	0	0	0	0	0
	Word order	0	0	0	0	0	0
	Article	0	0	0	0	0	0
	Preposition	0	0	0	0	0	0
	Total	**0**	**1**	**6**	**1**	**1**	**0**
Discourse	Coherence and cohesion	0	0	1	0	0	0
	Paragraphs	0	0	0	0	0	0
	Sylistics	0	4	6	0	0	0
	Total	**0**	**4**	**7**	**0**	**0**	**0**
Other	Other	**0**	**0**	**0**	**0**	**0**	**0**
Total		**6**	**26**	**45**	**3**	**7**	**6**

Appendix 35: Frequencies and proportions of the resolution of episodes on the basis of categories

GROUPS	RESOLUTION																							
	CONTENT				MECHANICS				LEXIS				FORM				DISCOURSE				OTHER			
	C[1]	I[2]	A[3]	IG[4]	C	I	A	IG	C	I	A	IG	C	I	A	IG	C	I	A	IG	C	I	A	IG
CYCLE 1																								
STAGE 1																								
CG	12	0	0	0	26	16	0	0	20	12	7	7	35	10	0	1	5	1	0	0	2	0	0	0
%	8%	0%	0%	0%	17%	10%	0%	0%	13%	8%	5%	5%	23%	6%	0%	1%	3%	1%	0%	0%	1%	0%	0%	0%
TG	23	1	2	1	29	8	0	0	39	13	6	4	36	14	0	0	9	0	0	0	0	0	0	1
%	12%	1%	1%	1%	16%	4%	0%	0%	21%	7%	3%	2%	19%	8%	0%	0%	5%	0%	0%	0%	0%	0%	0%	1%
LTG	15	2	0	0	37	9	1	0	13	14	6	7	48	9	0	0	11	0	0	0	1	0	0	0
%	9%	1%	0%	0%	21%	5%	1%	0%	8%	8%	3%	4%	28%	5%	0%	0%	6%	0%	0%	0%	1%	0%	0%	0%
STAGE 2																								
CG	2	0	0	0	7	4	1	0	2	3	1	0	15	7	0	0	7	2	0	0	0	0	0	0
%	4%	0%	0%	0%	14%	8%	2%	0%	4%	6%	2%	0%	29%	14%	0%	0%	14%	4%	0%	0%	0%	0%	0%	0%
CYCLE 2																								
STAGE 1																								
CG	2	2	0	2	42	11	0	0	17	6	4	1	32	7	0	0	3	0	0	0	4	0	0	0
%	2%	2%	0%	2%	32%	8%	0%	0%	13%	5%	3%	1%	24%	5%	0%	0%	2%	0%	0%	0%	3%	0%	0%	0%
TG	13	3	0	0	26	9	0	0	13	19	4	1	23	13	0	0	11	0	0	0	0	2	0	0
%	9%	2%	0%	0%	19%	7%	0%	0%	9%	14%	3%	1%	17%	9%	0%	0%	8%	0%	0%	0%	0%	1%	0%	0%
LTG	3	0	0	0	17	4	0	0	13	6	3	0	24	1	0	1	14	0	0	0	1	1	0	0
%	3%	0%	0%	0%	19%	5%	0%	0%	15%	7%	3%	0%	27%	1%	0%	1%	16%	0%	0%	0%	1%	1%	0%	0%
CYCLE 2																								
CG	8	0	2	1	11	9	0	0	3	1	1	0	11	3	0	0	5	0	1	0	0	2	0	1
%	14%	0%	3%	2%	19%	15%	0%	0%	5%	2%	2%	0%	19%	5%	0%	0%	8%	0%	2%	0%	0%	3%	0%	2%

[1]Correct; [2]Incorrect; [3]Addressed; [4]Ignored