UNIVERSIDAD DE MURCIA

ESCUELA INTERNACIONAL DE DOCTORADO

TESIS DOCTORAL

The Effects of Written Corrective Feedback Processing Conditions on L2 Learners' Written Texts and Levels of Depth of Processing

Los Efectos de las Condiciones de Procesamiento de Retroalimentación Escrita en los Textos Escritos por Aprendices de un Segundo Idioma y en sus Niveles de Procesamiento

D. Sophie McBride

2023

UNIVERSIDAD DE MURCIA

ESCUELA INTERNACIONAL DE DOCTORADO

TESIS DOCTORAL

The Effects of Written Corrective Feedback Processing Conditions on L2 Learners' Written Texts and Levels of Depth of Processing

Los Efectos de las Condiciones de Procesamiento de Retroalimentación Escrita en los Textos Escritos por Aprendices de un Segundo Idioma y en sus Niveles de Procesamiento

Autora: Da. Sophie McBride

Directora: Da. Rosa María Manchón Ruiz

UNIVERSIDAD DE MURCIA

DECLARACIÓN DE AUTORÍA Y ORIGINALIDAD
DE LA TESIS PRESENTADA PARA OBTENER EL TÍTULO DE DOCTOR

Aprobado por la Comisión General de Doctorado el 19-10-2022

D./Dña. Sophie McBride

doctorando del Programa de Doctorado en

869- Programa de Doctorado en Artes y Humanidades: Bellas Artes, Literature, Teologia, Traducción e Interpretación y Lingüística General e Inglesa (Plan 2013)

de la Escuela Internacional de Doctorado de la Universidad Murcia, como autor/a de la tesis presentada para la obtención del título de Doctor y titulada:

The Effects of Written Corrective Feedback Processing Conditions on L2 Learners' Written Texts and Levels of Depth of Processing.

Los Efectos de las Condiciones de Procesamiento de Retroalimentación Escrita en los Textos Escritos por Aprendices de un Segundo Idioma y en sus Niveles de Procesamiento

y dirigida por,

D./Dña. Rosa María Manchón Ruiz

DECLARO QUE:

La tesis es una obra original que no infringe los derechos de propiedad intelectual ni los derechos de propiedad industrial u otros, de acuerdo con el ordenamiento jurídico vigente, en particular, la Ley de Propiedad Intelectual (R.D. legislativo 1/1996, de 12 de abril, por el que se aprueba el texto refundido de la Ley de Propiedad Intelectual, modificado por la Ley 2/2019, de 1 de marzo, regularizando, aclarando y armonizando las disposiciones legales vigentes sobre la materia), en particular, las disposiciones referidas al derecho de cita, cuando se han utilizado sus resultados o publicaciones.

Si la tesis hubiera sido autorizada como tesis por compendio de publicaciones o incluyese 1 o 2 publicaciones (como prevé el artículo 29.8 del reglamento), declarar que cuenta con:

- *La aceptación por escrito de los coautores de las publicaciones de que el doctorando las presente como parte de la tesis.*
- *En su caso, la renuncia por escrito de los coautores no doctores de dichos trabajos a presentarlos como parte de otras tesis doctorales en la Universidad de Murcia o en cualquier otra universidad.*

Del mismo modo, asumo ante la Universidad cualquier responsabilidad que pudiera derivarse de la autoría o falta de originalidad del contenido de la tesis presentada, en caso de plagio, de conformidad con el ordenamiento jurídico vigente.

En Murcia, a 14 de marzo de 2023

Fdo.:

Esta DECLARACIÓN DE AUTORÍA Y ORIGINALIDAD debe ser insertada en la primera página de la tesis presentada para la obtención del título de Doctor.

Información básica sobre protección de sus datos personales aportados	
Responsable:	Universidad de Murcia. Avenida teniente Flomesta, 5. Edificio de la Convalecencia. 30003; Murcia. Delegado de Protección de Datos: dpd@um.es
Legitimación:	La Universidad de Murcia se encuentra legitimada para el tratamiento de sus datos por ser necesario para el cumplimiento de una obligación legal aplicable al responsable del tratamiento. art. 6.1.c) del Reglamento General de Protección de Datos
Finalidad:	Gestionar su declaración de autoría y originalidad
Destinatarios:	No se prevén comunicaciones de datos
Derechos:	Los interesados pueden ejercer sus derechos de acceso, rectificación, cancelación, oposición, limitación del tratamiento, olvido y portabilidad a través del procedimiento establecido a tal efecto en el Registro Electrónico o mediante la presentación de la correspondiente solicitud en las Oficinas de Asistencia en Materia de Registro de la Universidad de Murcia

For you, Tessa.

ACKNOWLEDGMENTS

My journey towards PhD completion could not have been done without the support of the following people, to whom I owe the world. I was going to keep this short and sweet, but each and every one of you deserves a whole thesis, so here goes...

Thanks to my family and friends back home for always being there, despite the distance. We don't get to see each other often, but I know you are all just a phone call away. To my parents, who worked endlessly to provide the best opportunities for me. To Ben, Natalie, Frankie and George, thanks for the laughs, I love you. To Aunt Diane, you are incredible and your daily WhatsApp messages of encouragement have meant the world to me. To Jo, Bazza, and the rest of the McBride clan, thank you!

Nanna, I have dedicated this thesis to you, words cannot describe how much I adore you and how much I have missed you over the past few years whilst working on this project. Thank you for always supporting me, I love you.

Lucy, Aimee, and Sophie. I literally have no words to describe how much you mean to me, I absolutely adore you and I want to thank you for your patience and understanding whilst I have been working on this thesis. Giovanna, Steph and Jess, the ultimate squad! Thanks for the banter, you cannot imagine how much it has helped!

To my Murcian family, where do I even begin? Words on this page will not suffice to express my gratitude for the immense encouragement you have given me throughout the past few years. Aida, Ana, Andrea, Cristina, Lola and Marta, you are all one in a million and you have provided me with a home and a family and for that I will be eternally grateful.

Mu, you deserve a special mention, your emotional support, the home-cooked meals, the dog-sitting for Kili whilst I travelled to conferences (thanks to Tita Lola, too!), there aren't enough words to thank you for everything you and your family have done for me, and I can only hope that someday I can repay you. I genuinely could not have done this without you.

To my fellow “Precarios”. Sharing this experience with you and knowing you understand every single feeling has been a heart-send. Thank you for the laughs, the coffees, the therapy sessions, and the encouragement to never give up. Annie, Corpus, Fernando, Juampe, Julia and Rodri, my wonderful group of misfits, you are amazing!

I would also like to thank some other people who over the years have demonstrated a kindness towards me that I will forever be in debt for.

Esperanza, Carlos, Carolina, and Pablo, my extended Murcian family, I love you all.

Antonio, little Antonio, Maria Jose and Marina, the weekly visits gave me a rest bite from thesis stress, and I cannot thank you enough for all your help and encouragement throughout the years.

Jenni, thank you for providing me with a sanctuary to escape to when the city became too much, you have the purest soul, and I am so grateful to call you my friend.

A special thanks to Kristin, one of the strongest, most hard-working individuals I have ever had the pleasure of working with. Your advice has been invaluable, and I have learnt so much from you both academically and personally. Thank you also to Marge, your strength is like no other and I just know you are going to achieve great things, you deserve no less.

Aitor, thank you for all your help and for always answering my WhatsApp audios rambling on about stats and writing. You are an absolute star! To Maria Dolores, Alberto, Belen, Joaquin, Arturo, Julio, Yvette, Flori, Luke, Raquel, Sonia, and Lena, you have helped me grow so much academically and I cannot thank you enough for the emotional support, I could not have done this without you.

J.A., although you are no longer a member of the linguistics clan, you (and your music) have helped me so much over the past few years, thanks my lad, you are a superstar!

A special thanks to Ana Pellicer and Marije Michel for allowing me to carry out my research stays. I will be forever grateful for the kindness you showed me whilst I was away from home.

Also, a special thank you to Ronald Leow, for treating my so kindly at conferences and always pushing me to ask more questions! You are a true academic role model.

Lourdes, my mentor, friend, and adopted family member. Praise the day I last-minute decided to write my undergraduate dissertation on Applied Linguistics instead of English literature! If it weren't for my spontaneous (and somewhat ridiculous) actions, we would have never met! You have gone above and beyond for me, not only providing me with incredible academic support, but you have also given me a place to call home. Thank you for always looking out for me and allowing me to share some wonderful memories with you, Fran and Alonso. I really hope to make you proud one day!

Rosa, my mentor, and my role model. It has been an absolute privilege to work with someone so talented, so successful, so ambitious and yet so kind. You have consistently thrown me in the deep end and pushed me to work towards my dreams, providing me with the most incredible opportunities I could have ever asked for. Thank you for encouraging me to be the best version of myself, both in and out of the classroom.

Thank you to my students, past, present, and future, I adore teaching and my weekly classes were a much-needed rest bite from my investigative duties.

To my participants, I could not have completed this project without your help, thank you from the bottom of my heart for your hard work and dedication.

I would not have been able to complete this dissertation without the funding I received as part of my pre-doctoral fellowship and as a a member of the L2 writing research group financed by the Spanish Research Agency (Research Grant PID2019-104353GB-I00 funded by MCIN/AEI/10.13039/501100011033 and Pre-Doctoral Grant BES-2017-081873) as well as The Seneca Foundation (Research Grant 20832/PI/18).

Last but by no means least, may the adventures continue, until the sky falls down on me...

THE EFFECTS OF WRITTEN CORRECTIVE FEEDBACK PROCESSING CONDITIONS ON L2 LEARNERS' WRITTEN TEXTS AND LEVELS OF DEPTH OF PROCESSING

Sophie McBride

Thesis Supervisor: Rosa María Manchón Ruiz.

ABSTRACT

The present PhD attempted to contribute to two relevant SLA-oriented lines of research: (i) effects of composing medium on written texts; and effects of feedback processing conditions on writing processes and products. The intended contribution was empirical, including a central methodological aim. The motivation for these global aims derives from the following considerations. On the one hand, in response to the mass shift seen in language classrooms to more online, digital learning environments, L2 writing scholars have advocated for research to explore and compare the effects of traditional pen-and-paper versus digital composing environments on the cognitive processes involved in writing and feedback processing, as well as on the characteristics of the resulting written texts (e.g., Vasylets & Marín, 2022; Vasylets et al., 2022; Zhi & Huang, 2021). On the other hand, SLA-oriented research on written corrective feedback (WCF) has received ample attention throughout the years (as recently reviewed by Bitchener, 2021; Roca de Larios & Coyle, 2021; Hyland & Hyland, 2018; Kang & Han, 2015, 2021), with studies that focus specifically on the processing of WCF gaining increased attention more recently. The construct of depth of processing (DoP, Leow, 2015, 2020) of WCF has become a key concern in theoretical and empirical feedback research. Studies in this domain have employed diverse methodological procedures -including think-aloud protocols (e.g., Bowles & Gastañaga, 2022: Caras, 2019; Kim & Bowles, 2019; Leow et al. 2022; Sachs & Polio, 2007) and written languaging (e.g., Cerezo et al, 2019; Manchón et al, 2020; Suzuki, 2012, 2017) to obtain data on potential (i) effects of DoP on how deeply L2 users engage with the feedback provided on their writing, and (ii) correlations between DoP and language acquisition (usually operationalized in terms of improvements in text revisions). Importantly, scholarly debates have more recently focused on research methodological considerations regarding data elicitation procedures in this research. In this regard, some critics (e.g., Leow & Manchón, 2021; Manchón, 2023a) have called for more controlled, methodologically oriented studies in which the validity of the data collection instruments is tested, whilst also advocating for investigations in more diverse writing environments.

In response to these calls, the main aims of the present doctoral thesis were to contribute empirically to previous research by exploring writing and feedback processing in both pen-and-paper and digital environments, and to shed light on the affordances of diverse introspective measures (individually and combined) commonly used for WCF processing. To achieve these global aims, the following research questions guided our study:

> **RQ.1** How does writing in a traditional pen-and-paper environment versus writing in a computer-mediated environment affect L2 written production in terms of CAF measures?

RQ. 2 How does the experimental manipulation during WCF processing affect L2 written production (in terms of CAF measures) in pen-and-paper versus computer-mediated writing environments?

RQ.3. How does the experimental manipulation during WCF processing affect L2 learners' levels of depth of processing of the feedback received in pen-and-paper versus computer-mediated writing environments?

To answer RQ2 and RQ3, the study explored the methodological affordances of three WCF processing conditions: (i) think-aloud protocols, (ii) written languaging, and (iii) a combination of think-aloud protocols and written languaging in two writing environments (computer-mediated and pen-and-paper writing conditions).

The study followed a pre-test/treatment/post-test design in which 36 English undergraduate students participated. Participants were invited to write an initial text (pre-test) in time-constrained conditions. The writing task was the problem-solving, picture-based "Fire Chief" task (Gilabert, 2007), which was completed by half of the participants (18) online, via GoogleDocs, and by the remaining 18 participants on pen-and-paper. Regardless of the writing and processing conditions, all participants received unfocused, direct WCF on their initial written texts. The participants were then invited back to process the feedback received, according to the treatment group to which they were assigned: (i) think-aloud only, (ii) written languaging only, and (iii) simultaneous think-aloud and written languaging. The final task (post-test) invited participants back to rewrite their original text under the same conditions as in the pre-test. Once the processing data had been collected, the think-aloud protocols were transcribed and coded following Leow's (2015) definition of DoP and the written languaging data was coded according to the levels of engagement and noticing, guided by the coding scheme elaborated in Cerezo et al. (2019). The written products were analysed in terms of a range of CAF measures.

Results show that, as regards composing medium, computer-mediated written texts were initially found to be more accurate and more fluent when compared to more traditional pen-and-paper written texts. Additionally, composing medium played a role on how the participants engaged with feedback: the pen-and-paper condition was more successful in engaging students in metalinguistic languaging, which, as a result, led to higher levels of accuracy in subsequent revised texts. In terms of feedback processing conditions, results show that the combination of think aloud and written languaging whilst processing WCF was not only the processing condition that provided the most insights into WCF processing, but it also constituted the most favourable treatment condition for promoting higher levels of DoP. These deeper levels of processing also led to higher L2 accuracy in subsequent text revisions.

These results represent relevant, novel insights into writing and feedback processing in diverse writing environments, while at the same time the insights obtained point to equally relevant and novel methodological implications for future research by shedding light on the affordances of WCF processing instruments and conditions. In addition, and from a pedagogical perspective, the results also present a series of potential implications relevant to pedagogical decision-making in the second language classroom with regards to writing task implementation and feedback processing conditions.

LOS EFECTOS DE LAS CONDICIONES DE PROCESAMIENTO DE RETROALIMENTACION ESCRITA EN LOS TEXTOS ESCRITOS POR APRENDICES DE SEGUNDAS LENGUAS Y EN SUS NIVELES DE PROCESAMIENTO

Sophie McBride

Directora de tesis: Rosa María Manchón Ruiz.

RESUMEN

La presente tesis intenta contribuir a dos líneas de investigación relevantes dentro de los estudios de segundas lenguas (SLA): (i) efectos del medio de composición en los textos escritos; y (ii) efectos de las condiciones de procesamiento de la retroalimentación en los procesos y productos de escritura. Este trabajo persigue objetivos tanto de naturaleza empírica como de índole metodológica, cuya justificación deriva de una serie de premisas. Por un lado, los estudios de escritura en segundas lenguas (L2) han abogado principalmente por una investigación que explora y compara los efectos de los entornos de composición más tradicionales en papel con los digitales en los procesos cognitivos implicados en la escritura y en el procesamiento de la retroalimentación, así como en las características de los textos escritos resultantes (por ejemplo, Vasylets & Marín, 2022; Vasylets, Mellado & Plonsky, 2022; Zhi & Huang, 2021). Por otro lado, la investigación enmarcada en estudios de SLA sobre el feedback correctivo escrito (WCF) ha recibido una amplia atención a lo largo de los años (como han revisado recientemente Bitchener, 2021; Roca de Larios & Coyle, 2021; Hyland & Hyland, 2018; Kang & Han, 2015, 2021), siendo los estudios que se centran específicamente en el procesamiento del WCF los que han ganado mayor atención más recientemente.

El constructo de profundidad de procesamiento ("Depth of processing", DoP, Leow, 2015, 2020) del WCF se ha convertido en una preocupación clave en la investigación teórica y empírica de la retroalimentación. La profundidad de procesamiento se define como la cantidad relativa de esfuerzo cognitiva y el nivel de análisis, junto con el uso de conocimientos previos, comprobación de hipótesis y formulación de reglas empleados en la descodificación y codificación de algún elemento gramatical o léxico (Leow, 2015), y sirve para medir el grado de procesamiento cognitivo que realiza un alumno cuando aprende un nuevo elemento lingüístico. Aplicado a los estudios de retroalimentación, permite medir el grado de implicación de alumnos cuando reciben feedback, así como arrojar luz sobre las acciones en las que se involucran los sujetos cuando reciben retroalimentación escrita.

Los estudios en este ámbito han empleado diversos procedimientos metodológicos, incluyendo protocolos de pensamiento en voz alta (por ejemplo, Bowles & Gastañaga, 2022: Caras, 2019; Kim & Bowles, 2019; Leow et al. 2022; Sachs & Polio, 2007) y reflexión sobre la lengua por escrito ("written languaging", ejemplo, Cerezo et al, 2019; Manchón et al, 2020; Suzuki, 2012, 2017). Con estos dos procedimientos, se obtienen datos sobre los posibles (i) efectos de condiciones experimentales de procesamiento en la profundidad de procesamiento del WCF recibido, y (ii) correlaciones entre DoP y adquisición de la L2 (generalmente operacionalizada en términos de mejoras en la revisión de textos). En general, los resultados de la investigación existente apuntan a beneficios en la implementación de actividades de procesamiento de feedback y, a pesar de numerosas diferencias en las metodologías empleadas en los estudios, el hecho de procesar retroalimentación escrita conlleva a mejoras en la revisión de textos escritos.

Es importante también destacar que los debates académicos se han centrado más recientemente en las consideraciones metodológicas relativas a los procedimientos de obtención de datos en esta investigación. A este respecto, algunos críticos (por ejemplo, Leow & Manchón, 2021; Manchón, 2023a) han abogado por estudios controlados guiados por metas metodológicas en los que se compruebe la validez de los instrumentos de recogida de datos, al tiempo que abogan por investigaciones en entornos de escritura más diversos. En respuesta a estas propuestas, la presente tesis doctoral pretende contribuir empíricamente a la investigación mediante el análisis de las características de textos y del procesamiento de la retroalimentación tanto en entornos de escritura tradicionales en papel y digitales, así como arrojar luz sobre las posibilidades ofrecidas por diversas técnicas de introspección (consideradas individualmente y combinadas entre ellas) comúnmente utilizadas en la investigación sobre procesamiento de WCF.

Para lograr estos objetivos globales, las siguientes preguntas de investigación guiaron nuestro estudio:

> **RQ.1** ¿Cómo afecta el entorno de escritura en papel y en entornos digitales a las características de los textos escritos en términos de corrección, fluidez y complejidad?
> **RQ. 2** ¿Cómo afecta la manipulación experimental durante el procesamiento de WCF a la producción escrita de L2 (en términos de medidas de corrección, fluidez y complejidad) en entornos de escritura en papel y en entornos digitales?
> **RQ.3**. ¿Cómo afecta la manipulación experimental durante el procesamiento de WCF a los niveles de profundidad de procesamiento en entornos de escritura en papel y en entornos digitales?

Para responder a las RQ2 y RQ3, implementamos tres condiciones experimentales de procesamiento: (i) protocolos de pensamiento en voz alta; (ii) reflexión escrita, y (iii) una combinación de ambos. El estudio siguió un diseño pre-/post-test en el que participaron 36 estudiantes universitarios de inglés como L2. Se invitó a los participantes a escribir un texto inicial (pre-test) en condiciones de tiempo limitado (50 minutos en total). La tarea de escritura correspondía a la "Fire Chief Task" (Gilabert, 2007), que consiste en una tarea en el que los alumnos deben proporcionar soluciones a un problema (en este caso, los participantes deben salvar a un numero determinado de personas de un edificio en llamas). La tarea fue completada por la mitad de los participantes (18) en línea, a través de *GoogleDocs*, y por los 18 participantes restantes en condiciones de escritura en papel. Independientemente de las condiciones de escritura y de procesamiento, todos los participantes recibieron WCF global y directo sobre los textos escritos como pre-test, es decir, todos los errores encontrados dentro del texto fueron marcados y la corrección correspondiente fue aportada a los participantes.

A continuación, se invitó a los participantes a procesar la retroalimentación recibida, según el grupo de tratamiento al que hubieran sido asignados: (i) pensamiento en voz alta, (ii) reflexión escrita, y (iii) simultáneamente pensamiento en voz alta y reflexión escrita. El tratamiento de pensamiento en voz alta consistía en invitar al participante a verbalizar todo lo que se le pasaba por su mente mientras procesaba el feedback escrito directo. Los alumnos realizaron esta fase del estudio individualmente y sus verbalizaciones fueron grabadas. La tabla de reflexión consistía en una actividad escrita en la que los participantes tenían que anotar los errores que habían cometido, indicar el tipo de error, anotar la corrección proporcionada, y dar una explicación metalingüística del error. Los participantes pertenecientes al grupo de tratamiento simultaneo tuvieron que realizar las dos actividades a la vez, es decir, rellenar la tabla de

reflexión escrita, mientras verbalizaban sus pensamientos en voz alta. La tarea final (postest) consistió en la re-escritura del texto original en las mismas condiciones que en el pretest.Para ello, los participantes recibieron una copia de su texto original (sin correcciones) y se les pidió lo volviesen a escribir corrigendo cualquier error que considerasen necesario.

Una vez recogidos los datos de procesamiento, los protocolos de pensamiento en voz alta se transcribieron y se codificaron siguiendo la definición de DoP de Leow (2015), en cuatro niveles de procesamiento: (i) alto, (ii) medio (iii) bajo y (iv) nulo. Los datos de las tablas de reflexión escrita se codificaron sobre la base del esquema de codificación elaborado en Cerezo et al. (2019). Por tanto, los datos de introspección escrita se dividieron en cinco sub-niveles que correspondían a tres grados de detección ("noticing"), (i) detección a nivel de reconocimiento (nivel 1), (ii) detección a nivel de información (niveles 2 y 3), y (iii) detección a nivel de comprensión (niveles 4 y 5). Una vez analizados los datos de introspección individualmente, se creó un esquema de codificación común para poder comparar y contrastar los resultados hallados en ambas condiciones de tratamiento (McBride & Manchón, 2023).

Los textos escritos se analizaron de acuerdo con una serie de medidas de corrección, fluidez y complejidad (sintáctica y léxica). La corrección se midió tomando el numero total de errores lingüísticos, dividido por el numero total de palabras, multiplicado por 100. La fluidez de los textos se calculó teniendo en cuenta el numero de palabras escritas por minuto. Por ultimo, la complejidad de los textos se calculó utilizando una herramienta online que mide la complejidad léxica y sintáctica de los textos escritos (Synlex, Lu, 2010).

Los principales resultados del estudio son los siguientes. En lo que respecta al medio de composición, los resultados muestran que los textos iniciales escritos en ordenador fueron más correctos que los textos escritos en papel, resultado que podemos atribuir a las herramientas de corrección ortográfica integrada en *GoogleDocs.* Además, los textos en ordenador también demostraban una mayor fluidez, gracias al carácter facilitador del teclado, frente a la lentitud del uso de bolígrafo en papel. En cuanto a las medidas de complejidad léxica y sintáctica, se encontraron diferencias mínimas entre los dos entornos de escritura (escritura en digital y escritura en papel).

Los datos muestran que el entorno de escritura también influyó en el modo en que los participantes procesaron la retroalimentación recibida. La condición de escritura en papel tuvo como resultado una mayor actividad metalingüística por parte de los participantes, lo que a su vez condujo a mayores niveles de corrección en los textos escritos tras el procesamiento de la retroalimentación. En cambio, la condición de escritura digital llevó a los alumnos a procesar la retroalimentación recibida de una manera mas personal (siendo mucho menos metalingüístico en comparación de la condición de escritura en papel), consiguiendo así unos niveles más bajos de procesamiento.

En cuanto a las condiciones experimentales de procesamiento de la retroalimentación, los resultados muestran que la combinación de pensamiento en voz alta y reflexión escrita durante el procesamiento de la retroalimentación no sólo fue la condición de procesamiento que proporcionó más información sobre el procesamiento del WCF de los participantes, sino que también resultó ser el tratamiento más favorable para promover mayores niveles de profundidad de procesamiento, así como una mayor corrección en textos escritos como post-test. Sin embargo, si examinamos los instrumentos de introspección individualmente, cada uno de ellos conlleva una serie de ventajas que no se pueden descartar. Por un lado, las tablas de reflexión escrita sirvieron como una guía durante el procesamiento de la retroalimentación ya que garantizaron que

los participantes trataran todos y cada uno de los errores, puesto que tenían que anotarlos en la tabla. Por otro lado, la actividad de pensamiento en voz alta, aunque no tan eficaz para conseguir que los participantes detectaran todos los errores, desde el punto de vista metodológico, este instrumento resultó muy ventajoso para arrojar luz sobre los procesos cognitivos de los participantes mientras procesaron la retroalimentación escrita, ya que proporcionó datos de introspección, algo no aplicable a las tablas de reflexión, cuyos datos Daban cuenta del resultado del procesamiento, si bien no del procesamiento en sí mismo

A pesar de los avances empíricos y metodológicos que los resultados de la tesis puedan suponer, la investigación realizada tiene una serie de limitaciones que deben tenerse en cuenta. En primer lugar, los participantes en el estudio contaban con sólida formación lingüística y un nivel avanzado del inglés como L2. Ello tiene implicaciones para la generalización de resultados a otros aprendices de niveles inferiores de dominio de la L2 y de formación académica distinta. Por tanto, sería relevante explorar los efectos de instrumentos de introspección de entornos de escritura (digital y tradicional) ampliando el espectro de aprendices de L2, prestando especial atención a contar con representación de diversos grados de dominio de la L2 y diversa formación en lengua y lingüística, algo especialmente relevante al tener en cuenta el alto grado de procesamiento lingüístico que conlleva el procesamiento de la retroalimentación escrita.

Otra limitación de nuestro trabajo tiene que ver con el hecho de que empleamos sólo un tipo de tarea (una tarea de resolución de problemas compleja), con un solo tipo de retroalimentación escrita (directo), hechos que de nuevo limitan la generalización de resultados. Como consecuencia, sería relevante estudiar diversos tipos de feedback, así como diferentes tipos de tareas para poder discernir de forma más precisa los efectos del entorno de escritura y de las condiciones experimentales del procesamiento de la retroalimentación cuando se realizan tareas de diverso tipo y se procesa distintos tipos de retroalimentación. Por ultimo, debido a la dificultad que supuso la pandemia para la recogida de datos, solo contamos con un grupo reducido de participantes, quienes realizaron el estudio en un tiempo limitado. La investigación futura se beneficiaria de un estudio de corte longitudinal, con un mayor numero de participantes, con el fin de ver si los efectos observados en esta tesis se corresponden con ganancias de aprendizaje a largo plazo.

A pesar de las limitaciones mencionadas, los resultados representan nuevas y relevantes perspectivas sobre la escritura en L2 y la retroalimentación en diversos entornos de escritura. Los resultados obtenidos apuntan a implicaciones metodológicas relevantes y novedosas para futuras investigaciones al arrojar luz sobre las posibilidades de los instrumentos y condiciones de procesamiento de la retroalimentación. Además, desde una perspectiva pedagógica, los resultados presentan también una serie de implicaciones relevantes para la toma de decisiones pedagógicas en el aula de segundas lenguas con respecto a las condiciones de realización de tareas de escritura y de procesamiento de la retroalimentación sobre los propios textos.

Table of Contents

LIST OF FIGURES

LIST OF TABLES

CHAPTER I. INTRODUCTION

The concept of error is somewhat controversial in second language acquisition (SLA) research. There are divided opinions when it comes to defining errors and even more so when discussing whether or not errors should be corrected. However, from a pedagogical point of view, language learners of all ages undoubtedly receive feedback on their errors, and, at some point, there will come a time when they are faced with processing these errors, in order to fully comprehend where they went wrong and how they can improve.

Some see errors as a natural, developmental process. In the same way a baby learns through making mistakes, an adult second language learner may also benefit from this development. However, if we take a more pedagogical stance and view errors as detrimental, the correction of errors seems like the most logical and beneficial next step in language learning. Thus, despite the contended view on errors and the treatment they should or should not receive, this thesis delves into error making, error correction and feedback processing and it does so from an SLA perspective. Additionally, in response to the mass shift that language classrooms have experimented towards more digital, online environments, the thesis also aims to contribute to L2 writing research on the effects of traditional pen-and-paper versus digital writing on both the characteristics of written texts and feedback processing.

In addition to shedding light on the effects of writing environments on writing and WCF processing, the research conducted also attempted to contribute to methodological debates on the instruments implemented in WCF processing research by exploring the affordances of two commonly used introspective measures, namely, written languaging tables and think-aloud protocols, as well as the combined used.

This dissertation consists of two main parts, which comprise a total of seven chapters. The first part of the thesis (Part I) provides the background to the study. Chapter 2 explores theoretical and empirical research on second language writing and written corrective feedback, looking at studies that have included digital and pen-and-paper writing, as well as research that has specifically looked into WCF processing, by exploring both oral and written introspective measures.

Part two includes chapters 3 to 7 and corresponds to the study that was carried out for the doctoral thesis. Chapter 3 details the aims of the thesis and the three research questions that guided the study. Chapter 4 provides a detailed description of the

methodology of the study that was carried out, detailing the data collection procedures followed, the coding schemes adopted for the data collected, and data analyses procedures. Chapter 5 presents the quantitative and qualitative results for each of the three research questions guiding the research. Chapter 6 provides a detailed discussion of the results, again divided according to the three research questions. Finally, Chapter 7 provides the conclusions, limitations, and implications of the current study, and provides suggestions for future research agendas.

PART I.

BACKGROUND TO THE STUDY

CHAPTER. II. LITERATURE REVIEW. THE LANGUAGE LEARNING POTENTIAL OF WRITING AND WCF

II.1. WRITING TO LEARN

The concept of writing to learn emerged primarily from the late seventies through to the early eighties when several articles were published claiming the extensive qualities writing exercises possess in order to enhance student learning, with a primary focus on L1 learning in a variety of school-based subjects. Amongst them, Emig's seminal work *Writing as a mode of learning* (1977) described writing as representing "a unique mode of learning" which constitutes a "powerful learning strategy" (p. 122). She described writing as a tool for not only learning content in particular subjects but also as an instrument capable of promoting language specific learning. This idea by Emig preceded James Britton's book *Prospect and Retrospect. Selected Essays* (1982), which described writing as a device that:

> [C]an in fact be learning in the sense of discovery. But if we are to allow this to happen, we must give more credit than we often do to the process of shaping at the point of utterance and not inhibit the kind of discovery that can take place by insisting children know exactly what they are going to say before they come to say it. (1982, p. 110)

In his chapter on *Writing to Learn and Learning to Write*, Britton explores the learning properties writing can have for children's language development, referring specifically to *expressive writing,* which relates to writing as a means to explore new ideas. As stated by Bazerman et al. (2005, p. 57), this specific type of writing "could play a cardinal role in learning at every developmental stage, in part because it resembled what Vygotsky had identified as *inner speech*", (discussed further in section 2.4.4. of this doctoral thesis). As a result of these two landmark publications, a growing body of research amassed, focusing on the potential learning qualities of writing, and providing data that demonstrated the complex yet beneficial nature of the writing process as a whole (Applebee, 1984). A summary of the main findings in the literature up until this point can be found in Applebee (1984) in which he states three primary findings in writing to learn research:

(1) Writing does not consist in a linear sequence but involves repeated subprocesses such as planning, monitoring, drafting, revising, and editing;
(2) All writers vary in their use of these processes;
(3) The writing task itself influences the processes a writer uses.

In relation to finding (1), in particular, the perdurability of writing facilitates the processes mentioned. The tangibility of written work allows writers to review and change their written texts, accordingly, moving back and forth as they write, unlike in oral settings (Applebee, 1984; Emig, 1977). However, and as stated in finding (2), not all learners work in the same way and their approach to writing processes may depend on a series of factors. Extant research has been carried out in the field of individual differences (IDs), an area which explores issues such as motivation, anxiety, learning goals, and language aptitude and explores the ways in which these IDs may affect learners writing and consequent learning outcomes (for an overview on ID research in L2 writing, see Kormos, 2012, 2023). Finally, and in addition to learner variables, the very task itself plays a crucial role in how a learner approaches the writing process. Variables such as task genre, time on task, and task complexity all play an important role in how a writer produces a text and have been a widely investigated area, particularly in L2 writing research (an overview on task-variables and L2 writing can be found in Byrnes & Manchón, 2014 and Robinson, 2011).

What was indisputable throughout the literature (Applebee, 1984; Britton, 1982; Emig, 1977) and later empirically corroborated (e.g., Langer and Applebee, 1986; Newell, 1984), was that regardless of individuals approaching writing tasks in various manners (in terms of their writing processes), the very act of writing itself holds the power to lead to learning.

II.1.1. Writing and Language Learning

Despite the surge of research focusing on the learning potential of writing in an L1 setting (primarily concerning the writing across the curriculum [WAC] movement throughout the 70s in the United States), it was not until fairly recently that research on writing in the field of second language acquisition (SLA) gained pivotal attention, with the scholar Alistair Cumming among the first to consider the potential of L2 writing as a means for language learning. In his pioneering work, Cumming (1990) stated that

"writing elicits an attention to form-meaning relations that may prompt learners to refine their linguistic knowledge" (p. 483). What had previously been dismissed as a point of interest in SLA investigations, in favour of research on oral production, eventually gained a new perspective and writing as a site for language learning has more recently become a key interest in SLA research. This regard for writing was propelled a few years later in an influential paper by Linda Harklau ("The role of writing in classroom second language acquisition", 2002) published in the *Journal of Second Language Writing*. After collecting data in L2 classrooms, it became evident to Harklau that what was primarily seen in SLA research as the main source of language learning: oral input and production, was actually absent in the classrooms she was observing. What she did find however, was the presence of writing (and reading) and as a crucial factor in the learning process of the students under observation, so much so, that some of the individuals included in the data collection reported that they would tend to "tune out" and ignore spoken input in favour of focusing on written work (p. 331). As a result and added to "the lack of attention to the role of literacy in classroom language learning" (p. 332), Harklau made a stand for more research to be carried out on writing as a tool for language learning, echoing what Cumming had voiced in 1990. Thus, what followed, was an increase in research that placed writing as a central role for L2 language acquisition. One initial influential work was a chapter by Manchón and Roca de Larios (2007). In their contribution to the volume on *Intercultural Language Use and Language Learning*, Manchón and Larios discussed the psycholinguistic rationale for exploring the language learning potential of writing in an L2 environment and also provided a synthesis of the empirical work that had been carried out in this area, looking specifically at the "problem-solving" nature of L2 writing (p. 104). Manchón and Roca de Larios' programme of research contributed a plethora of data to the field of SLA in which behaviours concerning the problem-solving nature of language as well as the many instances of linguistic processing involved in writing could be observed. More precisely, they contributed new evidence on the crucial role writing plays as a language learning tool in the L2 classroom by shedding light on the writing processes of native Spanish speakers. Results showed that the most crucial element involved in this process was time, therefore implying that the temporal dimension that writing affords (as opposed to the ephemeral nature of the oral modality), is essential for engaging learners in the writing task and crucial for the "problem-solving behaviour while writing that is deemed to be conducive to learning" (Manchón & Roca de Larios, 2007, p. 117). The authors called for future research to empirically test the theoretical

predictions that had been made thus far in the field, and also expand research on L2 writing by looking into task-related variables, and written corrective feedback provision (Manchón, 2020).

On account of these three influential works addressing the importance of writing as a tool for language learning and the call all three contributions made for more research on writing in SLA, Manchón published in 2011 an edited volume dedicated to the intersection between L2 writing and second language acquisition titled *Learning-to-Write and Writing-to-Learn in an Additional Language*. In this title, the importance of viewing writing not only as a skill to be learnt in the L2 classroom but also as a potential site through which second or foreign language could be learnt was asserted. Manchón called for future research to focus more on "the language learning potential [LLP] of written output practice, including both the act of writing and the processing of feedback on one's own writing" (2011, p. 77). Consequently, since the publication of this edited collection, research in the field has explored numerous lines of inquiries on writing and written corrective feedback studies, with both theoretical and empirical contributions adding to the existing research. As a result, numerous edited collections have been published on both writing (e.g., Byrnes & Manchón, 2014; Manchón, 2012; Manchón & Matsuda, 2016; Manchón & Polio, 2022; Storch, 2013) and written corrective feedback (e.g., Bitchener & Ferris, 2012; Bitchener & Storch, 2016), with Manchón's 2020 book *Writing and Language Learning* exploring the interface between L2 writing and L2 WCF. In addition, journals in the field of SLA have also dedicated special issues to L2 writing and WCF including for example, the 2021 landmark special issue in the *Journal of Second Language Writing* (Exploring L2 writing-SLA interfaces*,* 21) which collected a range of papers addressing the disciplinary debates stated in Manchón (2011)*.* Other special issues include: "The Role of Oral and Written Corrective Feedback in SLA" *(Studies in Second Language Acquisition,* 32, 2010), "Methodological Advances in Investigating L2 Writing Processes" (*Studies in Second Language Acquisition*, 41, 2019), "L2 Writing and Feedback Processing and use in Pen and Paper and Digital Environments: Advancing Research and Practice" *(Studies in Second Language Learning and Teaching,* 12, 2022) and an upcoming special issue guest edited by Rosa M. Manchón and Cristina Sanz, "Working Memory and L2 Learning: Implications for Research on Individual Differences in SLA" *(Studies in Second Language Acquisition*, 2023).

As a result of the proposals to consider writing as a crucial site for second language learning (Cumming, 1990; Harklau, 2002; Manchón & Roca de Larios, 2007),

and the body of empirical research that followed exploring the effects of writing on L2 language gains (aforementioned edited collections and journal special issues), it became clear that the theories of L2 writing as a facilitator of language learning were verifiable. Therefore, research began to theorize why writing led to gains in the L2 and what were the specific characteristics writing held that enabled these gains (Manchón, 2023a). Three main characteristics attributed to writing were said to be key in its role as a site for language learning, including: (i) the off-line nature of writing, (ii) the permanence of writing and feedback, and (iii) the problem-solving nature of writing (Manchón & Cerezo, 2018).

The first condition relates to the pace with which writing is completed. As opposed to oral environments, writing allows for more extended time on task, which essentially provides learners with more time to reflect on their L2 knowledge, to attend to the input more carefully and likely notice more language-related issues. The second condition concerns the permanence of writing, in contrast to the ephemeral nature of oral communication. This visibility and permanence of the product allows learners to continuously focus their attention on the written output during the writing process and, if feedback is provided, return to their original written output, and engage in feedback processing by comparing their original version of the text, with the revisions they receive. The permanence of writing therefore provides extended opportunities for learners to notice the gap between their existing L2 knowledge and new target forms. The third characteristic of writing relates to the problem-solving nature of writing tasks. As stated in Manchón and Roca de Larios (2007), engaging in writing processes presents an opportunity for learners to tackle linguistic problems which need to be solved, and given the permanence and extended time writing presents, these language-related issues that arise can be solved during the composing stage or during the feedback processing stage. These three characteristics therefore present an optimal condition for language learning to take place and can be related to a number of SLA-oriented theoretical principles.

The following section will look in depth at the theoretical positions that have informed the empirical research carried out on the LLP of L2 writing, and which serve as a conceptual basis for the present doctoral thesis. This will be followed by a section analysing the empirical research that has been carried out on SLA-oriented L2 writing, with a particular focus, in line with the present thesis, on writing in digital and traditional pen-and-paper contexts.

II.1.2. Theoretical Underpinnings on Writing and SLA

In order to explore the L2 learning potential of writing and written corrective feedback, and before delving into the empirical studies that reflect this very potential, it is crucial to review the theoretical underpinnings that address the learning processes, specifically, through a second language acquisition lens. In particular, and in line with the doctoral thesis, theories relating to the key role played by attention and awareness in L2 writing and written corrective feedback processing, including cognitive theoretical frameworks such as Richard Schmidt's Noticing Hypothesis (1990), Merrill Swain's Output Hypothesis (2005), as well as DeKeyser's skill acquisition theory (2015) will be discussed. This is due to the fact that the language learning potential of L2 writing is premised on the role of attention in SLA. In addition, and relevant to the feedback processing stage of the thesis, a sociocultural lens will be included, looking specifically at Lev Vygotsky's (SCT) sociocultural theory (1978), which situates learning within interactions, mediated by others (teachers or peers) and oneself (Cumming, 2020).

II.1.2.1. Schmidt's Noticing Hypothesis

One of the very first theoretical works to directly address the importance of attention and input in the field of SLA was the Noticing Hypothesis (1990 and elsewhere). Richard Schmidt discussed the role of attention from a cognitive perspective and focused his hypothesis on the early stages of learning, that is, the input stage. This theory can be directly related to the feedback stage of writing, relevant to this doctoral thesis, as it focuses on the crucial role of attention when receiving input in order to learn. According to Schmidt and based on his own personal experiences as an L2 learner, without attention, an L2 learner has minimal chances of transforming the input they receive to intake. What's more, in order for an L2 form to be learnt, as well as attention, Schmidt argued that a minimal level of awareness was also essential, which he defined as *noticing*. "I use "noticing" to mean conscious registration of the occurrence of some event" (Schmidt, 1995, p. 29). If a learner is able to notice an item in the L2, they are then likely to convert this into intake and therefore, learning may take place, thus "Schmidt rejects the idea of learning without awareness" (Leow, 2015). On a higher level, and although not necessary for learning according to Schmidt, is the concept of awareness at the level of understanding (Leow & Driver, 2021). This higher level of awareness relates to L2

learners hypothesising about the L2 input received, formulating rules concerning the form and comparing and contrasting this new input to their existing L2 knowledge. Ideally, a student will obtain this level of understanding when presented with input, as it implies that the target form has been learnt and stored in their long-term memory. This item learning can be achieved when a form is noticed (without understanding) but this implies learning at a lower "surface level" (Schmidt, 1995, p. 29) and therefore in order to achieve optimal learning, the more an L2 learner notices, the more they are likely to learn (Schmidt, 2001). If we situate this hypothesis within the current thesis, it is possible to view the crucial role noticing plays not only on writing but also on feedback processing as, by providing error corrections to learners, the opportunities for noticing to take place increase. In addition, the more explicit the feedback provided, the more likely a student will notice features in the input and, consequently, compare it with his/her existing knowledge. If the linguistic input has been analysed successfully, corresponding to higher levels of awareness (noticing with awareness at the level of understanding), it will likely lead to learning.

II.1.2.2. Swain's Output Hypothesis

In contrast to Schmidt's Noticing Hypothesis which focuses on the input to intake stage of the learning process, Swain's Output Hypothesis (2005) centres on the knowledge processing stage of learning, that is, the production stage (originally focused on oral production, this hypothesis is now also extensively applied to written production). For Swain, asking learners to produce language, raises their awareness of the possible gaps that exist between the L2 they are learning and their current L2 knowledge. If a student is able to recognise these gaps between their interlanguage and the new L2 forms, then there is a higher chance that learning will take place. As this hypothesis is situated within the production stage of the learning process, it is highly relevant to the production of L2 writing, the provision of WCF and, even more so, to the processing of feedback. Given the availability of written products over a prolonged period of time, in contrast to oral production, the potential for learners to hypothesise and test their L2 knowledge becomes greater, as learners are able to spend time writing and re-writing, whilst reflecting on both internal (self-evaluation) and external feedback processes (generally, teacher-led) and *noticing* the gaps between their existing L2 knowledge and what they

need/want to convey. Manchón and Roca de Larios (2007) summarise these ideas magnificently:

> Learners can pay "attention" to language while writing, and […] this is beneficial for language learning because (i) attention allows the L2 learner to become aware of the gap or mismatch between what they can communicate and what they would like to communicate, and (ii) through external feedback and the process of monitoring one's own productions, learners may also notice the gap between the rules underlying their production and the L2 rules, as well as the gap between what they can produce and what more proficient users of the L2 […] produce or ask them to produce. (p. 108)

If in addition to solely providing written corrective feedback, learners are requested to process this feedback, either orally or in written form, the language learning potential of writing is enhanced as the act of feedback processing encourages learners to "notice the gap" in their L2 knowledge mentioned in Swain (2005) and Manchón and Roca de Larios (2007). In addition to the function of (i) noticing triggered by a learners L2 production, Swain also discusses two other roles that output can play in L2 learning: (ii) hypothesis-testing, which allows learners to experiment with target forms and make changes accordingly (upon writing or receiving feedback, for example) and (iii) metalinguistic reflection which involves learners using language to talk about the language they have produced (metatalk). This last function relates directly to Vygotsky's sociocultural theory (explained below) and is crucial in understanding the rationale behind including written and oral languaging in the present thesis. By promoting learners to metatalk, student's ideas concerning language become "crystalised" and "inconsistencies [become] so clear […] opening the way for new learning to take place" (Swain, 2005, p. 479). In chapter II.3.3.3. an in-depth review of empirical research that focuses on the processing of WCF will be discussed and the concept of languaging as a potential language learning tool will be explored in depth.

II.1.2.3. DeKeyser's Skill Acquisition Theory

Another important theoretical position underpinning the connection between writing and language learning is DeKeyser's (2015) Skill Acquisition Theory (SAT). This

theory relates to the process a learner follows in L2 acquisition and emphasises the crucial role practice plays in the acquisition of L2 knowledge. The language learning process according to SAT, involves three predominant stages, which relate to the development of declarative knowledge, procedural knowledge, and automatic knowledge (terms originally coined by Anderson, 1982). The first stage, relating to declarative knowledge, describes the "explicit mental representation of language items" within a learner's L2 knowledge system (Iwashita & Dao, 2021, p. 283). This knowledge learners have *about* language must follow the process of proceduralization, in order to be fully learnt and incorporated into their L2 system. This process corresponds to the second stage of language learning according to SAT, in which the L2 learner must practice using the input in order to reach the third and final stage of ultimate attainment, corresponding to automatization. In order to develop proceduralised knowledge, a learner must engage in meaningful practice, which directly relates to the output stage of learning (Leow & Driver, 2021). Therefore, the very act of L2 writing and WCF provision yields learners with the opportunity to unerringly practice their declarative knowledge and potentially reach the later stages of acquisition in which terms become automatic, converting their knowledge *about* language into knowledge of *how to use* the language (Nassaji & Kartchava, 2021, p. 2)

Extrapolating this theory to L2 writing is feasible as the offline nature of writing provides the ideal setting for learners to practice their language and rely on their declarative knowledge in order to undertake the task. If additionally, corrective feedback is provided on the writing, opportunities arise not only for the proceduralization of existing knowledge but also for the acquisition of new declarative knowledge, as new L2 forms may be presented to learners via the provision of WCF. In terms of the feedback provided, indirect feedback may be more apposite in encouraging learners to self-correct and transition from declarative to procedural knowledge (Abbuhl, 2021, p. 48) whereas direct feedback, provides more opportunities for new L2 forms to be acquired and therefore, may lead to a higher chance of declarative acquisition. As SAT theory is based on the relevance of consistent practice of previously acquired declarative knowledge, it remains unclear to what extent feedback plays a role on all three stages postulated (Leow & Driver, 2021). Given that the most common feedback practice in language classrooms is the provision of one-shot feedback episodes, the introduction of new L2 items in this way does not necessarily provide learners with the opportunity to practice their newly acquired knowledge, therefore minimising the chances of proceduralization, according to

DeKeyser's theory (2015). However, if opportunities to process the feedback are provided to learners, via specific feedback processing tasks such as the ones included in this doctoral thesis (written languaging or think-aloud protocols), and opportunities are provided to practice the forms in subsequent rewritings of the text, proceduralization is more likely to take place.

II.1.2.4. Vygotsky's sociocultural theory (SCT)

The relevance of Liv Vygotsky's sociocultural theory in SLA research stems from the importance he placed on language as the principal source of knowledge and awareness (1978). For this scholar, "learning is a socio-cultural situation activity occurring through interaction with others […] via mediatory tools" (Moradian et al., 2020, p. 269), a learning process he refers to as the "inter-psychological place" (1978). For learning to successfully take place and become internalised, a transition from this *inter-plane* (learning through others) to an *intra-psychological plane* (learning autonomously) is essential. Thus, *others* are seen as a mediatory tool through which an individual receives help in order to acquire new knowledge. Language, of course, being the primary mode of interaction that leads to this desired construct of knowledge. As previously mentioned, (Swain, 2000), when learners are encouraged to use language to talk about language, albeit with mediators (other-regulated learning) or when engaging in metatalk alone (self-regulated learning), attention is addressed to new language forms, and potentially, understood, and finally, internalised.

It is important that the forms the learner is attempting to internalise remain within his/her *cognitive gap*. Looking again at the theories of Vygotsky, he spoke of the *zone of proximal development* (ZPD) when theorizing about the acquisition of new knowledge, defining the ZPD as, "the distance between the actual development 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" (1978, p. 86). In order for a learner to acquire a new form, not only does this form have to be within their reach developmentally speaking, but it is also essential that they are guided and encouraged by a peer who is more knowledgeable than them. If we extrapolate this theory to the provision of feedback in L2 writing, providing error corrections alone may not suffice for a learner to incorporate the new form into their L2 language system. We may hypothesise then that guidance may be needed in order for the

learner to truly take in the new information they are receiving from corrective feedback, following Vygotsky's postulations that this guidance should be provided by a peer or teacher who is "more capable" (p. 86), something which has been referred to widely in cognitive psychology research as assisted performance (Tharp & Gallimore, 1991) or *scaffolding*/scaffolded learning (Wood et al., 1976), later extrapolated and investigated in SLA literature (Bruner, 1983). One way of facilitating this essential guidance, particularly in a WCF setting, is by asking students to language (collaboratively or individually) about the errors and corrections they have received. Therefore, one of the aims of the present doctoral thesis, was to explore to what extent learners may benefit from the opportunity to language on their errors and to test ways (via different introspective measures) in which this learning process may be enhanced.

The following sections look in depth at the work that has been carried out thus far on L2 writing and written corrective feedback. First, the LLP of writing will be discussed, including the main research trends in the field. Then, writing modality will be discussed by comparing the research that has been carried out on the effects of writing in digital and traditional environments, an area of research especially relevant to the doctoral thesis.

II.2. RESEARCH ON WRITING AND LANGUAGE LEARNING: EMPIRICAL INSIGHTS AND OPEN QUESTIONS

The empirical research exploring the language learning potential (LLP) of writing can be divided into two main strands. First, research that has investigated the potential the act of writing itself has for language learning and, secondly, research that has focused on WCF as a tool that leads to L2 acquisition. The following sections will provide an overview of the research carried out on the LLP of writing, with a section devoted specifically to digital writing, given the focus of the doctoral thesis.

II.2.1. Empirical Research on the Language Learning Potential of Writing

Concerning the investigations carried out on the LLP of L2 writing, research can be divided according to a range of variables, with most studies focusing on (i) task-related factors; (ii) writing conditions; and to a lesser extent, (iii) writing processes (Manchón, 2020). In addition, and in response to Kormos's (2012) call for IDs to play a more pivotal role in L2 writing research, recent years have seen an increase in the importance of

individual differences (IDs) as moderating variables and thus, an ever-growing number of L2 writing studies have included both affective (e.g., Zabiji et al., 2020) and cognitive (e.g., Kormos & Trebits, 2012) IDs in their research design in order to analyse to what extent these variables may influence task outcomes and L2 written production.

In terms of task-related variables, issues concerning task complexity (e.g., Vasylets et al., 2019), task genre (Lu, 2011; Yoon & Polio, 2017), task modality (e.g., Zalbidea, 2017), task planning (e.g., Ellis & Yuan, 2004) and task repetition (e.g., Amiryousefi, 2016; Nitta & Baba, 2014) have received ample attention and research has explored intersections between task-based language learning (TBLL) and the LLP of writing. This body of research coincide in manipulating task conditions in order to explore the potential effects this has on language production, generally measured in terms of the complexity, accuracy, and fluency (CAF) of the written products. However, most research in the domain has led to inconsistencies in findings due to the divergent methodologies implemented in the research designs.

In L2 writing task complexity research, complexity of tasks has been found to influence L2 performance and language learning, primarily based on two models (designed to be applied to the oral mode). The first model, Skehan's Limited Capacity Model (2009) argues that the more complex a task, the more trade-off effects it has, forcing a learner to either produce more linguistically complex or more accurate L2 output, but both are not possible. In turn, the Cognition Hypothesis, (Robinson, 2001), states that the more complex a task, the more complex and accurate a learners L2 output (Vasylets & Gilabert, 2021). Research in the field of L2 writing (Vasylets et al., 2017; Zalbidea, 2017) has provided empirical evidence to support the cognition hypothesis whilst also demonstrating that task complexity most likely does not work in isolation as other mediating variables, including task modality, play a crucial role on task completion and appear to interact directly with cognitive task demands (Vasylets et al., 2020).

Task modality has been examined in L2 writing literature and research has provided empirical evidence on how task mode plays a role on the CAF measures of L2 learners. This research has compared both oral and written modalities and results have demonstrated that writing tasks appear to provoke a more complex output from L2 learners, when compared to oral tasks, with empirical studies reporting a higher accuracy in writing (e.g., Zalbidea, 2017). These results can be attributed to the aforementioned unique characteristics of writing which provide learners with a permanent, recursive resource in which they are able to focus on language-related issues without worrying

about the fleeting nature of oral communication. This, added to the availability of more time to complete a task, engages learners in a constant monitoring of their language production, leading to more complex language and higher accuracy in written forms (Vasylets et al., 2017; Zalbidea, 2017).

In relation to task repetition research, repeating certain writing tasks has been proven to enhance learner's attention to language forms and thus, lead to improvements in language accuracy (Manchón, 2014). The foundation of task repetition studies relates to the idea that allowing a learner to repeat a task provides them with the opportunity to divide their attentional resources between content (first iteration of the task) and language-related concerns (possible also in subsequent iteration) (Vasylets & Gilabert, 2021). Once again, the distinctive characteristics of writing provide learners with the opportunity to focus their attention on specific aspects of their texts, shifting back and forth between a focus on form and a focus on content, or between writing/revising different parts of the text, for example. This writing process has been considered as "internal task repetition" (Bygate, 2006; Manchón, 2014) and constitutes another fundamental benefit of writing tasks, which, added to the external task repetition nature of writing (via feedback provision and rewritings), enhances the possibility for language learning to take place. However, some researchers have suggested that the potential positive effects of task repetition may only play a significant role in writing done over longer periods of time and thus, have argued that these effects may only be truly visible in longitudinal studies (Nitta & Baba, 2014).

Another related line of enquiry in L2 writing research are the studies that look into the effects of task genre on linguistic complexity (Polio & Lee, 2017) which have found certain task types, such as argumentative texts, to be more conducive to complex language than others, such as narrative texts (Lu, 2011; Yoon & Polio, 2017). In terms of task planning, mixed results have been found for pre-text planning and while some research has found beneficial effects of planning on CAF measures (e.g., Ellis & Yuan, 2004; Farahani & Meraji, 2011; Lin, 2013), other results were not so clear (e.g., Johnson et al., 2012).

The discrepancies in results found in L2 writing and task-related research can be primarily attributed to differences in the methodologies used and, in most cases, to the measures implemented in rating L2 written production. Most research looked into similar writing tasks, with narrative being one of the most predominant types. However, many differences can be found in variables such as the participants involved in the studies, with

research including students from a wide range of L1 backgrounds and L2 proficiency levels, although, mostly adult learners are involved in the available research. Another main difference found is in the way L2 written production is measured: whereas some research favours holistic or analytic rating rubrics (particularly those studies investigating task-planning, e.g., Lin, 2013), others have used performance measures as manifested in the complexity, accuracy, and/or fluency (CAF) of the written product. However, one of the main issues leading to the discrepancy in results, relates to the different ways in which CAF dimensions are measured. Despite the abundance in research on task-related effects and due to the aforementioned discrepancies, more investigations into the intricacies of L2 writing and task-related variables are warranted, in order to "understand more fully the role of tasks in bringing about language learning through, by, and with writing" (Manchón, 2020, p. 417).

Another crucial strand that has been investigated in L2 writing relates to writing conditions, such as individual versus collaborative writing or, more recently, traditional versus digital writing. In terms of writing environment, despite the fact that research from other fields (e.g., psychology and neuroscience) has explored the effects of pen-and-paper versus computers on learning and performance (see review of this research in Vasylets et al., 2022), studies focusing on L2 written production in the two environments are scarce. The available research will be discussed in more depth in the following section, given the relevance this topic has for the present doctoral thesis. In relation to the empirical research carried out on collaborative (CW) versus individual writing conditions, studies have investigated CW in both traditional and computer-mediated domains and have compared these to individual writing contexts. CW conditions have proven to be effective for L2 writing development as well as beneficial for prompting interaction amongst students, leading to an enhanced number of language-related episodes (see Li & Zhang, 2021 for an overview of CW research in L2 writing).

Finally, the most recent line of investigation concerning the LLP of L2 writing corresponds to the actual process of writing and the manner in which engaging in L2 writing processes may contribute to L2 acquisition. This strand of research focuses on the different stages involved in the writing process including aspects such as planning, revising, and editing. In order to tap into these processes, researchers have implemented introspective measures such as think-aloud protocols and have analysed *decision-making episodes* (Cumming, 1990) and language-related episodes (López-Serrano et al., 2020) in an attempt to measure to what extent output production at the L2 writing process stage,

may lead to language learning. More recent studies have implemented digital tools such as keystroke logging, eye-tracking and screen capture (Garcés-Manzanera, 2022; Révész et al., 2019) to explore the cognitive processes writers engage in whilst composing texts, and to view aspects relating to macro-writing processes, such as planning, formulation, revising, etc.

II.2.2. Empirical Research on Digital vs. Traditional Writing

The L2 classroom has seen a recent shift towards blended and online learning as a result of the COVID-19 global pandemic, which has led to an increase in the incorporation of digital modalities in the L2 classroom. As a result, instructors have had to transition to using more digital tools and online methodologies in order to adapt to the evolving situation. Consequently, more traditional print-based modalities have been replaced by computer-mediated applications through which students and teachers are able to work collaboratively but more importantly, remotely. Focusing specifically on writing skills, this transition to online environments, has allowed instructors to implement a whole new range of innovative writing applications, specifically known as *Web 2.0 technologies.* Blog posts, *Wikis*, online forums and *GoogleDocs* are just some examples of the new digital genres that have been incorporated into L2 writing courses, which allow students to "share and contribute their thoughts/ideas with online communities by utilizing web-based software services [...] that encourage users to become more involved in the creation and manipulations of data" (Chang et al., 2012, p. 53). Recalling Vygotsky's sociocultural theory (1978) previously mentioned, and the importance of social interactions for language learning, Web 2.0 tools provide an innovative way of exposing students to authentic target language input, as well as providing crucial scaffolding assistance thanks to the facilitative nature of these tools to enable meditative interactions.

GoogleDocs is one of the applications that has been widely used in L2 classrooms as a result of the shift to online learning. This tool allows for synchronous, collaborative writing in which both students and instructors have access to the same document. As stated by Alharbi (2019), "this feature [...] maximizes its potential from allowing teachers and instructors not only to trace and facilitate student's work [...] but also to provide constructive feedback on student's work" (p. 3). In addition to the ease of receiving corrective feedback, *GoogleDocs* also allows learners to self-edit as they type, as with traditional pen-and-paper writing, the presence of the text in front of the learner, permits

them to constantly revise their work. Notwithstanding, digital writing is much more fluid in nature, allowing for the removal or addition of text with great ease, and at a much faster rate, allowing for longer texts in short periods of time (Goldberg et al., 2003). Added to this is the availability in many cases of tools such as *spellcheck* which provides suggestions to learners on improving their writing, particularly concerning punctuation, spelling mistakes, word order and tense use.

Most of the research available that focuses on digital writing in SLA, explores student's perceptions towards the implementation of computer-mediated writing, as opposed to more traditional pen-and-paper writing. In general, students react positively to digital writing, confirming a number of benefits including the clarity online texts provide, in terms of legibility (Vincent, 2016) and the convenience of *spellcheck* (Isaias, et al., 2015). Additionally, students have demonstrated a preference for digital writing when concerning more formal, academic texts (McBride & Garcés-Manzanera, 2022). In turn, regarding pen-and-paper writing, students have reported positive benefits for this modality particularly concerning its ability to lead to a higher knowledge retainment. Writing by hand is said to lead to a greater activation of a student's memory (Vincent, 2016) and is therefore preferred for activities such as notetaking and revising (McBride & Garcés-Manzanera, 2022).

Despite extant research available on student preferences, a gap in the SLA field exists when concerning empirical research that directly compares the L2 writing affordances of pen-and-paper versus digital writing, with most of the research comparing these two environments focusing more on reading skills. In other research fields including psychology and neuroscience, empirical evidence has attributed a number of advantages for both environments particularly concerning aspects related to brain activity, which sees an increase when participants are required to write by hand (Askvik et al., 2020), resonating with what many EFL participants also claimed (in the research on writing modality preferences) when attributing handwriting to a higher retention and therefore, higher learning of target items (McBride & Garcés, 2022; Vincent, 2016).

The research available in an L2 writing context has generally focused on writing tasks included in L2 writing tests, such as TOEFL (a critical review can be found in Cheung, 2012). Within this body of research, the L2 written products of test-takers have been analysed and the effects of paper-based writing as opposed to digital writing have been compared, resulting in a variety of findings. In a study by Wolfe, Bolton, Feltovich and Niday (1996) digital writing was found to be more beneficial in terms of text length

and the texts were found to be more formal in digital writing when compared to the texts composed on pen-and-paper. In line with these results, Whithaus, Harrison and Midyette (2008) found texts written on pen-and-paper were less formal and shorter in length. However, when concerning the quality of the texts in terms of the errors made, very minimal differences were found. In a more recent study by Chambers (2008) with low-proficiency learners, written texts produced in a digital modality were found to have a higher variation of lexical items (as measured by type-token ratio), which the author hypothesised could be due to the facilitative nature of digital writing for making revisions during the writing process. In terms of the quality of the written texts (and as found in Whithaus et al., 2008) findings were similar for both writing environments as both groups produced a similar number of errors in their texts. In line with previous research (Wolfe et al., 1996 and Whithaus et al., 2008), researchers Baraoui and Knouzi (2018) also found digitally written texts to be longer in length, with a higher variation and sophistication of lexical items (as found in Chambers, 2008). In addition to these findings, L2 proficiency played a role on the overall writing scores, regardless of writing mode, as those with a higher proficiency level, produced better texts than the lower proficiency level students. Importantly, a strong relationship was found between the participant's level of keyboarding skills and their overall writing scores, with students performing better on the writing task if they were more familiar with keyboarding. This mixed empirical evidence shows that writing environment may play an important role on the writing processes and writing performance quality of L2 learners (Vasylets et al., 2022).

To our knowledge, only one study has specifically focused on the effects of writing environment (pen-and-paper versus digital) on L2 writing that does not concern test-taking (Vasylets et al., 2022). An empirical study in which the authors also explored to what extent cognitive individual differences play a moderating role on any observed effects. In their initial analysis of the undergraduate student's writing, it was found that accuracy was higher for the participants in the digital writing group, with almost double the number of errors found in the texts written by hand. This finding can be attributed to the availability of spell check within the online writing tools used by the students (although the tools used by students are not explicitly specified in the study). In addition, fluency was found to be higher for the student's writing on pen-and-paper, a rather surprising finding considering the advantages keyboarding has for elaborating a text with greater speed and ease. However, the authors believe that the longer time spent on task for the digital groups, could be related to the fact that the participants spent more time

revising their work and thus, demonstrated a slower fluency. Revising in digital environments allows learners to make changes in their texts with ease, shifting back and forth within their texts more conveniently than on pen-and-paper. Due to this, learners have reported being more authentic in terms of their writing style when writing in digital environments as the inconvenience of revising on pen-and-paper leads them to "modify their natural writing behaviours" (Vasylets et al., 2022, p. 724) in an attempt to avoid too many revisions. In terms of complexity, no differences were found for the lexical items included in the texts, but a higher syntactic coordination was found in digital writing, confirming, as in previous research, that writing environment may play on role on the quality of written texts (Cheung, 2012). The findings suggest that the benefits the permanence of writing presents (as opposed to the ephemeral nature of oral language), may be enhanced in digital writing environments, as the permanence of the writing facilitates learners to constantly change the text with great ease (e.g., adding, deleting words, phrases or even paragraphs). As previously mentioned, one of the main conditions that constitutes writing as a site for language learning, is the opportunity of "internal task repetition" (Bygate, 2006), which in digital environments is made even easier. Thus, further research focusing specifically on L2 writing tasks, comparing the two writing environments (pen-and-paper versus digital) is crucial in order to shed light on exactly how much of a role environment plays on the quality of written texts, particularly in terms of the accuracy, fluency, and complexity of written products.

II.3. WRITTEN CORRECTIVE FEEDBACK AND LANGUAGE LEARNING

Written corrective feedback (WCF) has been defined by Leow (2020) as "any external manipulation of L2 writers' product by the teacher or the researcher designed to minimally draw their attention to some grammatical, lexical, structural, and/or content error committed by the L2 writers" (p. 99), forming an integral component of written production in the L2 classroom (Leow et al., 2022). The effectiveness of L2 WCF has been a focal point of research for many years, yet results have been conflicting, as discussed in the following section. Importantly, what is clear from the research is that, as with writing, WCF plays an important role in promoting L2 language acquisition. As previously mentioned, the premise on which writing and WCF are investigated, in terms of their potential for language learning, is centrally due to its permanent nature. For WCF, this permanence provides learners with the necessary time to reflect on their errors, notice the differences between their written output and the feedback provided, and subsequently

respond to the corrections provided when revising their original texts. In addition, the ability to manipulate the saliency of feedback (which can be enhanced by highlighting or underlining errors, for example), provides favourable conditions for learners to notice their errors, as opposed to what is possible when processing feedback in oral communication.

The following sections explore the LLP of L2 writing from a written corrective feedback perspective. Thus, we will first introduce the definition of WCF and an overview of the different types of feedback will be provided, including a summary of the main research trends in the WCF field. This will be followed by a discussion of the main theoretical positions informing WCF provision and processing in an SLA environment. After this, a review of the empirical research relevant to the doctoral thesis is presented in which the current investigations into WCF provision are provided, looking specifically at research on unfocused direct WCF as well as digital versus traditional WCF. Finally, and in line with the research carried out for the dissertation, the lens will be focused specifically on studies that look at WCF processing via oral and written introspective measures.

II.3.1. Written Corrective Feedback Definition and Types

An essential pedagogical practice in second language (L2) classrooms is the provision of feedback. Traditionally, this practice is provided by teachers/instructors who are expected, by their students, to engage in both oral and written observations that will promote a better understanding of the subject matter at hand and essentially, enhance L2 learner's oral and written production. What remains to be determined however is whether the provision of feedback is fully effective in facilitating second language acquisition (SLA) and if so, what types of feedback practices are best for enhancing this attainment, taking into consideration the innumerable factors surrounding a student's L2 learning potential. Research in the field, as we will see below, has tried to provide answers to these crucial questions.

Corrective feedback (CF) can firstly be defined according to the medium in which is it provided; therefore, two main types can be determined: oral and written CF. The main distinction between the two, concerns the way in which the information is processed. Due to the permanence of a written form of error correction, researchers including Manchón (2011), argue that written corrective feedback (WCF) may pose more benefits for L2 learners than oral feedback, especially in long-term effects, as it grants

learners a longer processing time in which they are able to compare their errors with the corrections provided. In an oral CF setting, the online nature may hinder a learner's chance for processing the error corrections and thus, curb any potential opportunities for uptake. In addition to the greater availability of WCF in terms of time, written error correction constitutes a direct, personalised observation that a learner can individually process and, hopefully notice. In turn, oral CF, again due to its online, fleeting nature, has the potential to go unnoticed by learner's who may simply not perceive that the CF is aimed for them, believing perhaps that it was directed at another learner.

The very first published articles concerning written corrective feedback (e.g., Truscott, 1996) held a rather negative perception towards the concept of WCF and the implementation of such methods in a classroom setting. It was argued that the provision of feedback to students could have a negative and detrimental effect on their overall learning (Truscott, 1996). This controversial assertion was responded to directly in an article by Ferris (1999) in which she disagreed with the claims and defended the positive effects that WCF may have in an L2 classroom. Stemming from these two articles, there has been a considerable amount of research on WCF in the L2 classroom, aiming at testing various forms of feedback provision, investigating which types may be more effective and confirming (or not) the positive benefits that WCF can provide students with. (See a summary of findings in Bitchener & Ferris, 2012; Bitchener & Storch, 2016; Kang & Han, 2015, 2021). As WCF research advanced, so did the focus on the investigations and what began as an investigation into whether WCF could be effective or not in an L2 classroom, progressively evolved into researching which types of WCF were most beneficial for learning (particularly in terms of the linguistic precision in their L2 written products), for whom and why.

Written Corrective Feedback (WCF) can be considered as any type of signalling of an error on either pen-and-paper or computer-mediated written texts. The form of indication used determines the type of WCF implemented. Therefore, in terms of the manner in which WCF is delivered, two methods have been categorized, namely, direct WCF and indirect WCF. In the cases when an error is indicated by a symbol, a metalinguistic code or simply by highlighting or underlining (depending on the degree of explicitness the instructor deems acceptable), the feedback is known as indirect WCF. This method of CF requires the L2 learner, not the teacher, to provide the correct L2 form. The very first studies in the field of L2 WCF leaned towards a preference for indirect WCF (Ashwell, 2000; Lalande, 1982; Polio et al., 1998), as they argued that this type of

WCF provides learners with a more cognitively demanding process, driving them to engage in language-related problem-solving and requiring them to think deeper about their errors, working towards self-editing, which is more likely to lead to long-term acquisition (Bitchener & Knoch, 2008; Ferris, 1995; Lalande; 1982; Park & Kim, 2019).

In turn, a body of research has focused on the provision of direct WCF (Bitchener, 2008; Bitchener & Knoch, 2009; Sheen, 2007), which consists not only in providing an indication of the error by the instructor/researcher, but also in supplying the correct L2 form. This type of feedback is supported by authors such as Chandler (2003), who argues that it can help students with more complex errors as they are provided with the correct linguistic form, which in some cases may not be within the students reach as they may not be linguistically ready, especially when it comes to beginner/low-intermediate learners. However, this type of feedback runs the risk of students simply copying the correct form and not truly understanding the meaning behind the correction, nor proceduralising the new information provided.

Research comparing both indirect and direct WCF have provided varying results. Despite early comparisons not finding significant differences between the two types of feedback (Frantzen, 1995), other studies have found benefits for indirect WCF (Ferris, 2006), whereas a range of studies have reported a clear advantage of direct forms of feedback (Bitchener & Ferris, 2012; Van Beuningen et al., 2012). These inconsistencies in the findings can be attributed to differences in study designs included in the research such as differences in task genres included, the length of treatment of the study, the quantity of writing required, additional L2 instruction, and feedback focus.

In addition to the classification of WCF in terms of the manner in which is it provided, WCF can also be categorised depending on the scope of the feedback. Focused WCF consists in only targeting the correction of one specific error type (Bitchener, 2008, Ellis et al., 2008). Instructors/researchers aim at fixing their attention on only one targeted item, providing learners with the opportunity to internalise these specific targeted forms. This would involve, for example, correcting all the errors in a given text related to the use of articles, or all the errors related to the use of verb tenses, etc. This type of correction can be observed in studies such as Bitchener (2008), Bitchener and Knoch, (2008, 2009, 2010), Ellis et al. (2008) and Sheen et al. (2009).

On the other hand, unfocused feedback (Ellis et al., 2008; Hartshorn et al., 2010; Nicolás-Conesa et al., 2019; Robb et al., 1986; Sheen et al., 2009; Truscott & Hsu, 2008, Van Beuningen et al., 2008; Van Beuningen et al., 2012) consists in providing feedback

that takes into consideration all of the errors made in a written text, correcting everything from grammar to punctuation to contextual mistakes. As with direct and indirect WCF, focused and unfocused WCF also have their own individual benefits and can provide language learning opportunities if used in the optimum conditions.

Empirical findings concerning these two types of WCF remain inconclusive. On the one hand, some studies were able to confirm that focused WCF was more effective than unfocused (see for example, Ellis et al., 2008). On the other hand, research has shown that unfocused WCF may have a positive effect on texts, especially in terms of their L2 accuracy (e.g., Van Beuningen et al., 2008). However, a range of methodological issues have been posited to have affected the results and therefore, some uncertainty remains as to how effective these two types of WCF may be for learners and also which type is most effective (Bitchener & Ferris, 2012).

A more focused approach towards WCF allows lower learners to pay closer attention to specific errors and provides them with the opportunity of noticing these errors and successfully understanding the corrections, without being overloaded with too much information at once. However, with learners of a higher proficiency level or learners who are perhaps paying for their language education, there's an expectancy of receiving WCF on all of the errors they have made and therefore, unfocused WCF not only responds to this demand but provides higher-proficiency learners with a chance to address all of the errors they may have made within a text. Recently, Liu and Brown (2015) in their meta-analysis on written corrective feedback mention a hybrid version of focused feedback, combining the two scopes, which have been found to be effective in the empirical research reviewed. They propose that rather than providing corrections on all errors or just one, a fusion of focused and unfocused ("mid-focused feedback") in which error corrections are provided on two to five error types "seems to be more practical and deserves more attention" (p. 74).

As mentioned however, the research carried out in the field of WCF and language learning has led to a range of contrasting findings, due to a large number of inconsistencies in the study designs. Some of the principal methodological issues in the research carried out on WCF types relate to the measure of learning the studies implemented. Much of the research has analysed results in one-shot studies, which mainly focus on short-term revisions of texts (Ashwell, 2000; Ferris, 2006; Ferris & Roberts, 2001; Truscott & Hsu, 2008). However, other studies have included a more longitudinal design in which they have compared the progress of students in terms of their linguistic

precision (or accuracy) in revised texts as well as in new writings (or delayed post-tests) with different intervals of time between the receiving and processing of the WCF (e.g., Chandler, 2003; Ferris, 2006; Ferris et al., 2013; Nicolás- Conesa et al., 2019). However, some scholars have criticised using text-revisions as a measure of L2 learning (Truscott & Hsu, 2008) and as a result, it was suggested (Bitchener & Ferris, 2012) that studies should include new texts in their research design, so as to truly measure language learning over time.

Despite the variety in focus, feedback type, the methods used, and language learning measures, research on WCF has been able to confirm that the opportunity to revise a written text can lead to improvements in the quality and the linguistic precision of the final written products (e.g., Bitchener et al., 2005; Shintani & Ellis, 2013; Shintani et al., 2014; Van Beuningen et al., 2012). However, it can be a major challenge deciding which type of feedback suits L2 learners best and both internal (e.g., learner's motivation, L2 proficiency) and external factors (e.g., curriculum, first language) should be taken into consideration when selecting WCF in the L2 classroom.

Another distinction made in the WCF literature, and one that has become more relevant in recent years, is the separation of research in traditional, pen-and-paper WCF on the one hand, and computer-mediated WCF, on the other. Due to the on-going increase in the appearance and use of computers in a classroom setting, accelerated by the recent pandemic which provoked a mass shift towards online learning, research has begun to include not only traditional provisions of WCF but also those environments where the error corrections provided are done so via a computer. Thus, studies have begun to question the potential effects computer-mediated WCF may have on a student's L2 language acquisition. However, the comparison between a more traditional, pen-and-paper setting with a modernised computer-mediated classroom remains a rather unexplored gap in the field and constitutes a rationale for the present study in which these two WCF environments will be explored empirically.

With the emergence of WCF studies that take into consideration new technologies in the classroom, comes the materialising of studies that concern feedback timing (e.g., Honeycut, 2001; Liu & Sadler, 2003; Ware, 2004). Throughout the past few decades, the timing of WCF has been a continuous progressing focus in WCF studies and researchers have aimed at exploring the effectiveness of providing immediate WCF as opposed to delayed WCF, attempting to shed light on the optimal timing of error correction. It is important to distinguish between the notions of immediate WCF as this can be interpreted

in a variety of ways. Firstly, immediate WCF can be perceived as an item-by-item feedback type in which participants are provided with error corrections after every item produced (this is typical of multiple-choice tests or language learning apps, for example). This particular type of feedback is salient in studies concerning synchronous and asynchronous computer-mediated WCF, in which synchronous WCF is generally provided item-by-item, in an online nature very similar to that of oral CF. Secondly, immediate corrective feedback can also be described as an end-of-test WCF in which a student receives corrections upon completing the task/sentence/essay. Again, this view of immediate correction provision is used both in oral and written CF, including in a computer-mediated environment. Thus, delayed corrective feedback can therefore be related to any type of feedback provided *a posteriori*, after the task has been carried out and, most usually, at least a day or two subsequent to task completion. A precise definition of concepts concerning the timing of WCF can be found in Lavolette (2014). For the purpose of this doctoral thesis, the empirical research review on digitally provided feedback that follows in section II.3.3.2. will focus specifically on asynchronous computer mediated WCF, as this was the type of digital WCF incorporated in the study.

Another extremely important distinction between feedback types when digital WCF is concerned is the "provider" of the feedback itself (Nurmukhamedov, 2009). Given the possibility a digital environment provides us with to include automated comments and corrections, computer-mediated feedback can be divided according to who, in effect, provides the corrections: (i) peer feedback, (ii) teacher-led feedback and (iii) Automated writing evaluation software (AWE). AWE corresponds to specific software that are implemented in order to provide scores on written texts, such as *Criterion* or *Grammarly.* This software works using algorithms and functions by comparing written work with a database so as to analyse features within the text relating to complexity, vocabulary, syntax, and total number of words (Hockly, 2019). As online learning evolves, so does the use of this type of feedback and therefore, the number of studies concerning AWE in an SLA context are increasing (see for example, Li et al., 2015; Zhang, 2020). Adjacent to this research is the work on peer feedback in a digital context, which has received abundant attention over the past few years (see for example, Tuzi, 2004). Conversely, research on teacher-led digital writing is less common and with the evolution of technology, less attention is being paid to teacher-led electronic feedback (see for example studies by Ducate & Arnold, 2012 and Eloloa & Oskoz, 2017). A more detailed review on computer-mediated WCF can be found in section II.3.3.2. in which

the empirical research carried out in this field, and relevant to the doctoral thesis, is analysed.

The overview provided of WCF in the field of L2 writing evidences the clear benefits the provision of WCF may have not only on written production, but also on language learning. However, in order to fully and truly understand why WCF can lead to language gains, it is essential to view the theoretical developments that have taken place in order to explore the language learning dimensions of WCF. The following sections review two important theoretical developments in the field, namely, two WCF and language learning models.

II.3.2. Theoretical Developments on Written Corrective Feedback and Language Learning

Despite the existence of a number of cognitive theories explaining L2 writing and second language acquisition, theoretical explanations in the area of WCF remain neglected and only recently have scholars begun to question how SLA theories can be attributed to the learning that takes place via feedback provision. The following section discusses two theoretical models that have been developed based on empirical findings in WCF research and also grounded in SLA theoretical approaches. These models are Leow's (2015) model of L2 learning processes in ISLA and Bitchener's (2016) model of WCF processing.

II.3.2.1. Leow's Framework for L2 learning Processes in ISLA

A recent theory on the language learning process was put forward by Ronald P. Leow in 2015, in which he included both processes and products in the framework and looked at the cognitive processes involved in all three stages of the learning process (input, intake and knowledge processing). Situating attention at the central role of the language processing model, Leow also introduces the concept of depth of processing as a crucial component of the L2 learning process. According to Leow, depth of processing can be defined as: "the relative amount of cognitive effort, level of analysis, elaboration of intake together with the usage of prior knowledge, hypothesis testing, and rule formation employed in decoding and encoding some grammatical or lexical item in the input" (2015, p. 204). Therefore, not only is attention crucial in the L2 learning process but depth of processing plays an essential role in various stages of L2 learning. Pertinent

to this doctoral thesis and its focus on WCF, the third processing stage (knowledge processing) is most relevant. As visible in the figure below by Leow (2020), the knowledge processing stage involves a loop in which any feedback received by the L2 learner is available to be processed further as new input (Leow, 2020, p. 103).

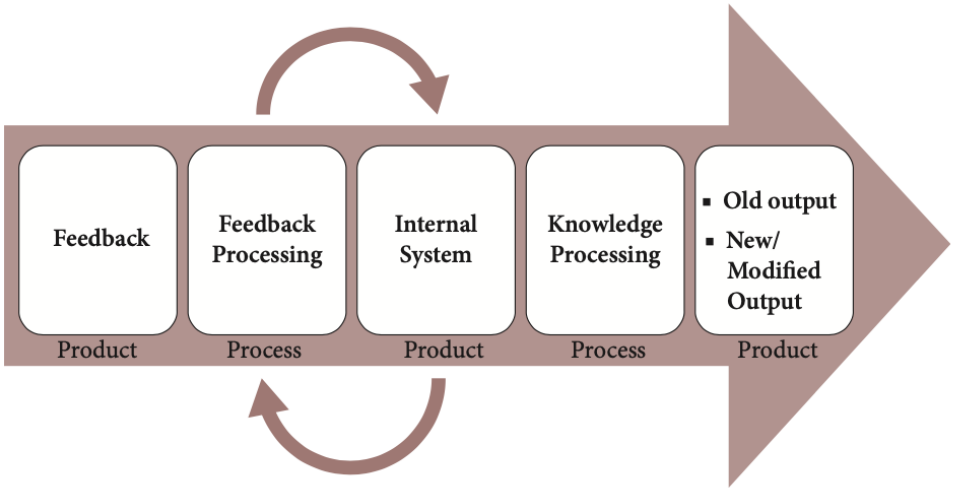


Figure 1. Feedback Processing Framework. Source: Leow (2020)

Consequently, this model is highly relevant not just for feedback research but also regarding the processing of WCF. In Leow's own words (2020, p. 104, emphasis in original)

> *Feedback Processing* encompasses *how* the learner cognitively processes the feedback (if at all) in relation to the current learner knowledge or interlanguage. If further processed at this stage, whether with a low or high depth of processing or level of awareness, the information in the feedback allows for reinforcement of accurate prior knowledge or, based on corrective feedback, for the potential of restructuring of previously learned inaccurate knowledge stored in the learner's *Internal System.*

By providing learners with the opportunity to process their errors upon receiving feedback, we are able to activate a process in which learners may take the new input received and transform it into new or modified output, if successfully processed (high

level of processing/awareness). Alternatively, if the new input is processed minimally (low depth of processing), there's a greater chance that this corrective feedback will not be incorporated successfully and thus, the learner will produce old/inaccurate forms in their future output. The end result of the knowledge processing stage is visible in the final output of the L2 learner and only then is it possible to view whether or not a specific item has been processed to a higher or lower level. As Leow states:

> Delayed performance may indicate whether a complete accurate restructuring took place [as in system learning] or whether such restructuring was temporary or immediate or reflective of item learning, that is, accurate performance was evidenced immediately after the feedback was provided but over time the learner reverted back to her previous inaccurate interlanguage (2020, p. 105).

By comparing pre- and post-tests in WCF studies, we are able to view just how much of the feedback provided has been successfully incorporated into new writings and, by zooming in specifically on the processing stage, to what extent depth of processing and awareness played a role in the acquisition of the correct forms.

II.3.2.2. Bitchener's Model of L2 Learning Processes

In an attempt to account for the actions learners take when processing WCF, Bitchener (2019) also proposed a framework to be considered for the provision and processing of WCF which was based on Gass' (1997) cognitive model of input processing. In this model for successful feedback provision, the author postulates that two pre-WCF provision conditions must be met *before* providing WCF (Bitchener, 2019). The first condition corresponds to the student's affective readiness, in the sense that students must be motivated to want to receive WCF. The second condition relates to the feedback focus, that is, the feedback provided must pay attention to form in order for learners to improve their L2 accuracy. If these two conditions are met, the WCF processing stage is more likely to be successful. Throughout the feedback processing stage, Bitchener (2019) postulates a sequence of stages relating to input processing, as seen in figure 2 below, which include: (i) attention to the WCF input, (ii) noticing the gap between the student's output and the instructors written input, (iii) an understanding of the WCF, (iv) an analysis or comparison of the WCF input and the learner's existing knowledge, and (v) hypothesis testing/formation (Manchón, 2023a).

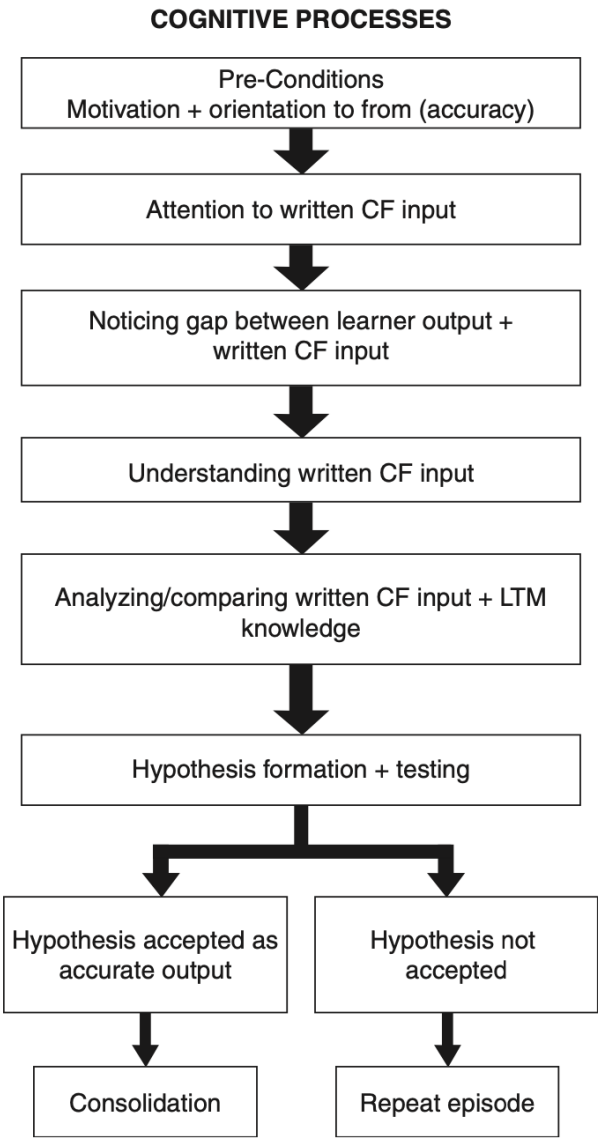


Figure 2. Cognitive Processing Stages during a WCF Processing Episode. Source: Bitchener (2019)

Not only does the model above represent the stages involved in processing WCF, but it can also be considered as a framework that represents instances in which students consolidate the knowledge provided from the feedback, in subsequent rewritings. As Bitchener (2019) states, "The same precondition identified [...] for the cognitive processing of a single episode of written CF would seem to be essential for learners seeking to access and use their new knowledge when writing new texts" (p. 92). Importantly however is the influence that moderating variables may have on the cognitive processes beforementioned. As previously discussed, the research carried out on WCF

has led to many contrasting results regarding the most advantageous feedback types and conditions. These inconsistencies can be attributed to moderating variables such as cognitive and affective individual differences and therefore, were included as moderating components in Bitchener's (2019) model of WCF processing, as shown in Figure 3.

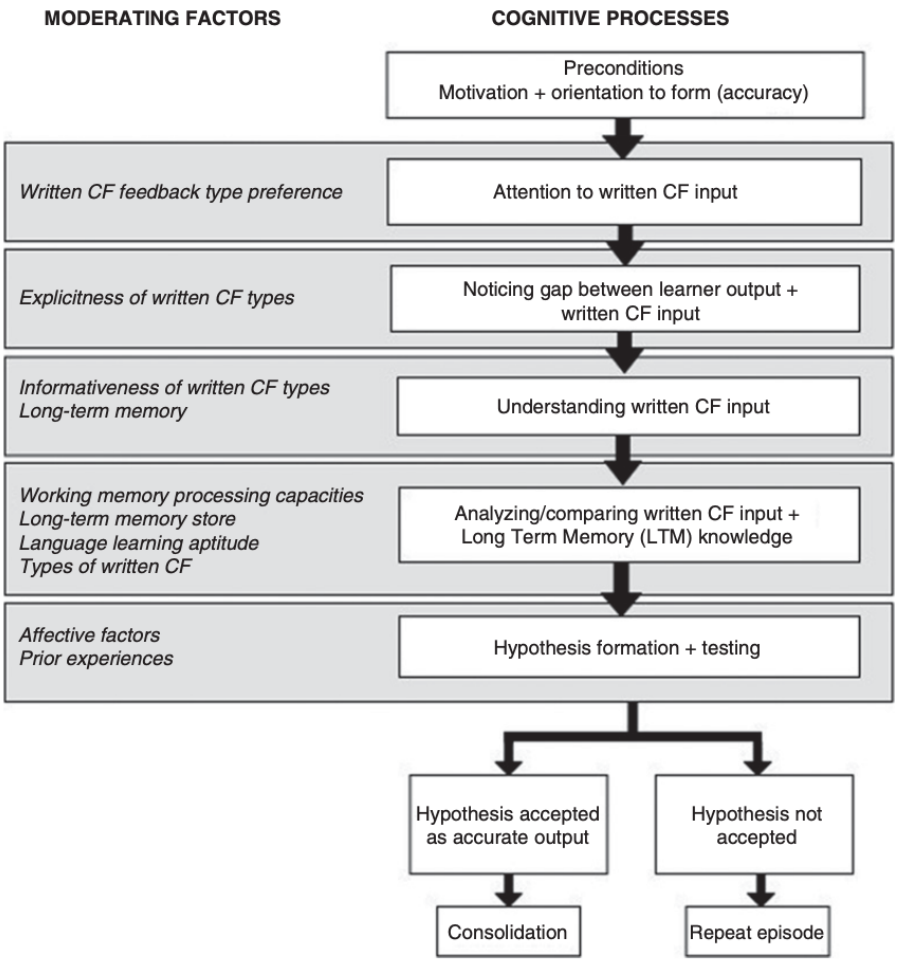


Figure 3. Model of WCF Processing. Source: Bitchener (2019)

Some factors contemplated in the model include external variables such as WCF explicitness and types, as well as leaner-internal variables such as working and long-term memory, language aptitude, prior experiences, and affective factors, including anxiety and motivation, for example (Bitchener, 2021). According to Bitchener (2019), rather than simply exploring how and why WCF is processed, it is essential to also explore the interactions of all moderating variables included in the model proposed.

Given the scarcity of WCF processing research, the two feedback models summarised above have yet to be tested in empirical studies and scholars have called for empirical evidence to be provided for the proposed frameworks (Manchón, 2023). The present doctoral thesis aims to contribute to the field of WCF processing, by exploring the ways in which learners in a higher education setting process the WCF they receive. As suggested by Manchón (2023), a first step in this needed model testing is to provide additional descriptive analysis of learner's engagement with WCF, exploring how learners make use of their prior knowledge, as well as how individual factors may influence their cognitive engagement with the feedback they are provided with (Manchón, 2023). This thesis contributes to partially filling these gaps.

II.3.3. Empirical Research on WCF in SLA

II.3.3.1. Unfocused Written Corrective Feedback in SLA

- *Unfocused Direct Written Corrective Feedback*

A common trend in the empirical research on feedback is a focus on the combination and comparison of a variety of feedback types. Therefore, there is scarce research looking at just one feedback type or focus. In Table 1 we see a summary of the primary research carried out on the provision of unfocused direct written corrective feedback. Both Kepner (1991) and Polio et al., (1998) carried out research in a university setting, with a similar number of participants. These studies were carried out over time and in the case of Kepner's (1991) study, participants in the first treatment group were provided with unfocused direct written corrective feedback on sentence-level errors as well as an explanation of the rules concerning these errors. In the second treatment group, participants were provided with content feedback. Thus, no control group was included. Results showed no improvements for these students over the course of the study in terms of their overall accuracy. However, these conclusions should be taken lightly as some flaws in the research should be highlighted. Firstly, Kepner (1991) did not include a control group in the study and additionally, no pre-test was provided to students in order to measure their initial accuracy levels. Interestingly, despite the feedback been described as ineffective for the participants in the results section, it has been argued that the data in

the study did indeed reflect a positive effect on the participants overall accuracy, as discussed in Bitchener & Knoch (2008).

Polio et al. (1998) also looked at accuracy improvements in their study and provided participants in the treatment group with unfocused direct WCF as well as instructions on editing for their text revisions. Participants were asked to work on a pre-test essay and were given the opportunity to revise this text upon receiving the unfocused direct WCF. Participants in the control group wrote four journal entries per week but did not receive any feedback on this writing. On the other hand, participants in the experimental group completed two journal entries and were asked to revise only one of these, in addition to completing a weekly grammar review and some exercises on the editing process. Participants were provided with feedback on the journal entries as well as the editing exercises. Results showed that both groups improved when it came to their linguistic accuracy and there was no greater improvement from the experimental group when compared to the control group. As the authors state in the study, these results may be more interesting from a pedagogical perspective rather than an empirical level as there is no way to measure how external factors may have influenced the results due to the longitudinally of the study (a full semester). This is of course something that is inevitable when collecting data in real classrooms and is always an additional factor that should be taken into consideration.

Table 1. Summary of Research on Unfocused Direct WCF

STUDY	PARTICIPANTS	DESIGN	TASK	TIME SPAN	FINDINGS
KEPNER (1991)	66 Spanish Intermediate University students	(1) Unfocused Direct WCF + explanation of rule (2) Content WCF	8 assignments (journal entries-no fewer than 200 words)	12 weeks	No improvements were found due to the WCF.
POLIO ET AL (1998)	64 University EAP learners	(Unfocused Direct WCF + instruction on editing (2) Control group	30 min essay writing + 60 min revision of essay	15 weeks	No significant improvement between experimental group and control group.

- *Unfocused Direct verses Unfocused Indirect Written Corrective Feedback*

In contrast to these studies looking at one type of feedback, a plethora of research has explored the effects of a variety of feedback types, including comparisons of the effects of unfocused indirect and direct WCF, with some studies including the additional variable of the availability of metalinguistic codes/information in their feedback provision. In 1986, Robb, Ross and Shortreed invited 134 first-year Japanese university students who were enrolled on an EFL composition course, to complete five narrative pieces of writing at home. The participants were divided into four experimental groups: an unfocused direct WCF group; an unfocused indirect WCF in which the errors were highlighted; an unfocused indirect WCF group in which the total number of errors were noted on the margins; and finally, an unfocused metalinguistic group in which the errors were coded. Results showed that all groups improved over time (one complete academic year), specifically in terms of syntactic complexity and fluency, regardless of the experimental condition they belonged to. Thus, the authors suggested that perhaps providing indirect unfocused feedback, which is considerably less time consuming, would suffice in the L2 classroom.

Almost ten years later, Frantzen (1995) published a study in which 44 university students of intermediate Spanish were divided into two experimental groups. This study varies significantly from others carried out including the same variables, as in addition to exploring the effects of unfocused direct verses indirect WCF, she also included the instructional environment as a variable. Therefore, the first experimental group consisted in a class focused on grammar in which the teacher provided students with unfocused direct WCF (with additional comments in the margins when clarification was needed) and the second experimental group consisted in a class focused on content in which the teacher provided students with unfocused indirect WCF, in the form of underlining or circles around errors made. Each participant was asked to complete four compositions in the classroom, of which the first and fourth were used as the pretest and the posttest, and a total of five, 250-word compositions at home over the course of a 15-week period. Results showed that both experimental groups improved their grammatical accuracy over time (measured according to the weighted grammar composite, corroborating with the results found previously in Robb et al., 1986). However, there were no significant improvements in their L2 fluency over the course of the university semester.

Moving now to a set of studies carried out in 2003, in a similar way to Robb et al., (1986), Chandler (2003) set out to respond to a gap in the research by looking at the combination of unfocused indirect WCF with unfocused direct WCF. In the second study, a total of 36 first- and second-year music undergraduates were divided into four experimental groups. Upon writing an autobiographical piece (average 8 pages long), participants were provided with unfocused feedback and a combination of responses including: (i) direct error correction, (ii) indirect corrections via underlining and description (metalinguistic codes), (iii) indirect corrections via description (metalinguistic codes) and finally, (iv) indirect corrections via underlining. These feedback types were provided to all students over the duration of the semester and Chandler wanted to explore whether or not students demonstrated an improvement in writing in terms of accuracy (calculated as the mean number of errors per 100 words) as well as fluency. Results showed that all groups improved in terms of their accuracy over the course of the semester. In addition, and opposing results found in Frantzen (1995), Chandler found improvements for fluency measures, as she found that students required less time when writing each subsequent chapter in the study.

These first three earlier studies comparing a variety of feedback types intended to add to existing literature by including more than one feedback variable. However, these studies excluded any control groups in the study which is a drawback and therefore, it cannot be assumed that the differences observed in the texts written before and after receiving feedback were due solely to the feedback received. That being said, these studies were carried out in classroom settings and, therefore, the inclusion of a control group would have been detrimental in terms of the pedagogical implications it would have had on this specific group, constituting an important ethical issue. Therefore, it is important to take into account the setting in which these studies took place and the justification behind the exclusion of such control groups.

More recent studies comparing unfocused feedback variations do include a control group in order to address this limitation. In addition, and as stated in two studies by Van Beuningen, Jong and Kuiken (2008, 2012), a further limitation that needed to be attended to was the lack of control with regard to the time spent on task. Thus, in the two studies, Van Beuningen et al., intended to contribute to the existing literature by exploring the effectiveness of both indirect and direct WCF on both short-term and long-term accuracy. In both of their studies (2008, 2012), in addition to the two unfocused WCF experimental groups (indirect and direct), two control groups were also included with one

group receiving extra writing practice opportunities while the other group self-corrected their errors, with no access to feedback. The two studies followed a similar method in which participants were asked to complete productive writing tasks in which they were invited to write an e-mail to a fellow pupil explaining the content they had seen in biology class, specifically, the metamorphosis of a variety of insects. In the 2008 study, 62 second-year Dutch secondary school students participated whereas in the 2012 study, a higher number of participants took part, with 124 higher level students from the second year of secondary education and 134 students belonging to the second year of prevocational education. Results from the 2008 study showed that both experimental groups (direct and indirect WCF) were effective in terms of accuracy on short-term writing. However, when looking at delayed post-tests and the long-term effects, only direct WCF resulted in significant accuracy effects. The control groups in the study showed no significant effects for accuracy improvement. Proving thus that comprehensive error correction can be extremely useful in order to enhance a student's accuracy. Interestingly, in the 2012 study, results showed that both direct and indirect CF contributed to an improvement in accuracy not only when concerning text revision but also in new pieces of writing over time (1 and 4 weeks respectively, after the feedback treatment session). Confirming again that the provision of comprehensive WCF constitutes an invaluable tool that can aid teachers in promoting and improving their student's L2 written accuracy over time.

Finally, looking at one of the most recent studies comparing unfocused direct and indirect WCF, Nicolás-Conesa, Manchón and Cerezo (2019) contributed new empirical evidence by exploring the effects of this feedback on rewritten texts as well as new texts. The participants in the study were 46 English undergraduate students from Spain who were divided into two experimental groups (receiving indirect or direct WCF) and one control group (who received no feedback). What differentiates this study from those previously mentioned is the inclusion of written languaging as a procedure for the students to engage in the processing of the feedback provided. Hence, both the control and experimental conditions engaged in an activity in which they were invited to reflect on the errors they had made and to fill out a table indicating the errors, the correction and a brief explanation. Results showed that those who engaged in written languaging benefited more from the feedback provision provided, particularly in terms of immediate uptake. However, the analyses revealed that a limited use of the feedback occurred in all experimental conditions, that is, only certain parts of the feedback provided was taken on

board by the participants. In line with previous research (Van Beuningen et al., 2008) they also found that direct WCF, in this case combined with written languaging, proved to be more effective when compared to indirect WCF for both re-written texts and the delayed post-test. The results concerning the specific effects of the written languaging treatment will be discussed in the next chapter which explicitly looks at empirical research on written corrective feedback processing. A summary of the main research carried out comparing unfocused direct and indirect WCF including the main research findings, can be seen in Table 2 below.

Table 2. Summary of Research on Unfocused Direct WCF versus Unfocused Indirect WCF

STUDY	PARTICIPANTS	DESIGN	TASK	TIME SPAN	FINDINGS
ROBB ET AL (1986)	134 Japanese first-year university students	1. Unfocused Direct WCF 2. Unfocused Indirect WCF group (metalinguistic codes) 3. Unfocused Indirect WCF (Highlighting) 4. Unfocused Indirect WCF group	5 Narrative test compositions	One academic year (23 classes-34,5 hours of classroom instruction)	1. Improvement was independent of WCF, no difference was found amongst experimental groups.
FRANTZEN (1995)	44 University learners of intermediate Spanish	1. Grammar class + Unfocused direct WCF 2. No grammar class + Unfocused indirect WCF (circling errors)	4 in-class compositions + 5 out of class 250-word compositions (memorable experiences)	15 weeks	1. Both groups improved their grammatical accuracy 2. Neither group improved their L2 fluency over time
CHANDLER-STUDY 2 (2003)	36 First- and second-year university students (Music)	1. Unfocused direct 2. Unfocused indirect + underlining + metalinguistic code 3. Unfocused indirect via explanation of error 4. Unfocused indirect via underlining only	Autobiographical writing (5 chapters-40 pages in total)	14 weeks	1. Improvements for both accuracy and fluency measures for all experimental groups. 2. Groups without metalinguistic information (1 & 4) performed better

					than those with codes (2 & 3)
VAN BEUNINGEN ET AL (2008)	62 second-year Dutch Secondary school students	1. Direct WCF group 2. Indirect WCF group 3. Control group (self-correction) 4. Control group (additional writing practice)	Two biology-related productive writing assignments (e-mails)	3 weeks	1. Short-term: both direct and indirect CF led to accuracy improvements. 2. Long-term: only direct CF was significant in terms of accuracy gains. 3. Neither control group had a significant effect on accuracy.
VAN BEUNINGEN ET AL (2012)	268 participants (134 higher level and 134 lower level) from four Dutch secondary schools	1. Direct WCF group 2. Indirect WCF group 3. Control group (self-correction) 4. Control group (additional writing practice)	Four biology-related productive writing assignments (e-mails)	6 weeks	1. Both indirect and direct CF led to improvements in both revision and new writing 2. Only Direct CF resulted in grammatical accuracy gains in new writing whereas indirect was more beneficial for nongrammatical accuracy.

NICOLÁS-CONESA ET AL (2019)	46 English undergraduates at a Spanish university	1. Unfocused Direct 2. Unfocused Indirect 3. No feedback	3 Narrative Tasks (personal experience)	5 weeks	1. Limited appropriation of WCF 2. Different errors benefit from different WCF types 3. Written Languaging promotes more benefits for learners for immediate uptake

II.3.3.2. Computer-Mediated Written Corrective Feedback

Traditionally, the research on corrective feedback in a written environment has primarily centred itself on a more traditional type of writing, that is, on pen-and-paper. Nevertheless, with the incorporation of new technologies in the classroom, there is a growing need to investigate electronic writing with new research on the effectiveness of various types of feedback in a digital writing environment. Available studies have focused on the role that AWE (Automated Writing Evaluation) may have in an L2 classroom (Gao & Ma, 2020; Lavolette et al., 2015; Li, Link & Hegelheimer, 2015; Stevenson & Phakati, 2014; Zhang & Hyland, 2017). These studies mainly conclude that automatic electronic corrections provided by applications in the classroom can lead to positive effects on student's L2 written production, particularly if combined with teacher-led feedback (Link et al., 2020). However, more profound investigation is needed in order to generalise the findings (see Mohsen, 2022 for a recent meta-analysis).

In addition to AWE, online applications such as *Google Docs, Microsoft Word* and *Microsoft's SkyDrive* allow teachers (and researchers) to provide their own feedback on writing. Teacher-led corrective feedback in a digital writing environment is commonly referred to as CMTF (computer-mediated teacher feedback) or electronic teacher feedback and has been defined as any type of feedback that is provided via the implementation of technological tools (Li, 2021). In a study by Elola and Oskoz (2016), *Microsoft Word* was implemented as the medium through which teacher-led feedback was provided to students and was compared with teacher-led oral feedback via *Screencast Software.* CMTF was found to be effective in not only engaging learners with the feedback but also in improving their L2 writing, proving once again that there are benefits for the integration of digital feedback resources in the L2 classroom.

More recently, and again using *Microsoft Word* for teacher-led feedback, Sarré, Grosbois and Brudermann (2019) explored the effectiveness of seven different feedback strategies in an online environment. 93 participants from the University of Sorbonne (Paris, France) were divided into seven groups; six experimental groups and one control group (receiving no feedback). The six feedback treatment conditions consisted in providing either focused/unfocused, direct/indirect and AWE/no additional AWE to the participants. Results from this study revealed that the groups who received CF on their writing outperformed the control group, thus, once again, confirming that any feedback is more beneficial than no feedback. Additionally, the authors found that the experimental group that performed the best

in terms of accuracy between the first and last writing task was the group who received unfocused, indirect WCF with additional computer-mediated "micro-tasks".

What is clear from the research carried out thus far is that there are certainly benefits to the implementation of teacher-led digital writing. However, the diversity in feedback types under study and the assortment of programmes and applications used to provide the feedback, make it extremely difficult to build well founded conclusions on the optimal feedback choices to be made in a teacher-led digital environment.

Along with the inclusion of new writing platforms in the classroom, comes the possibility for educators to simultaneously or asynchronously correct errors whilst a student/participant is producing a written text and thus it is highly relevant to introduce this style of WCF into current research as this trend in electronic writing is becoming widely used in the classroom. The ability to simultaneously provide WCF enables researchers to address the variable of timing in electronic WCF studies. It is now possible to compare the effects of synchronous (simultaneous) and asynchronous (delayed) feedback in order to view how timing may have an effect on the quality of a student's written production. In other words, what is the effectiveness of providing immediate corrective feedback, simultaneous to the writing process, instead of a delayed correction once the writing is finished. Studies such as the one carried out by Lavolette, Polio and Kahng (2015) investigated asynchronous WCF (1-3 weeks after the written task) and compared this with the provision of feedback after only a few minutes of elapsed time. The results did not demonstrate any statistical significance between the two types of feedback provision however and therefore, it was impossible to say which feedback timing was most beneficial for students/participants. Cerezo (2021), provides a synthesis of empirical studies in which spoken face-to-face (FTF) corrective feedback was compared with spoken/written synchronous computer mediated corrective feedback (CMCF). Cerezo concluded that although much evidence exists in favour of CMCF, in particular synchronous computer mediated communication (SCMC), the diversity in data collection procedures and variables included in the studies make it difficult to come to clear conclusions on which type of environment is best, especially for the research carried out on peer interaction, in which results are particularly contrasting (p. 511). As a result, and following Ortega's (2009, p. 245) call for more "studies pursuing a direct comparison between the SCMC and FTF modes", Cerezo (2021) further concluded that:

> It is everyone's responsibility to inform ourselves and others about how the different components of linguistic interaction (input, output, and corrective feedback) and

> multimodality (tools, learning objectives, and modes) operate and interconnect if we are to unlock the full potential of these communication opportunities (p. 514).

As with synchronous CMCF, research on asynchronous CMCF includes a diverse range of variables with most empirical studies focusing on peer provided feedback (Schultz, 2000; Tuzi, 2004), with less attention been paid to asynchronous teacher-led electronic feedback. In general, asynchronous digital WCF research can be grouped according to studies that focus on the comparison between electronic feedback and face-to-face CF (peer WCF: Schultz, 2000; Tuzi, 2004, and teacher-led multimodal WCF: Cunningham, 2019; Elola & Oskoz, 2016), research that focuses on the perceptions of students towards electronic WCF (Cunningham, 2019; Chong, 2019; Elola & Oskoz, 2016; Kiliçkaya, 2019; Tafazoli et al., 2014; Ware, 2014; Ware & O'Dowd, 2008), as well as research that looks into the focus of the feedback provided depending on the feedback type (Ene & Upton, 2014; Huang et al. 2020; Link et al., 2020; Ware & O'Dowd, 2008).

Despite the general lack of research on asynchronous digital WCF, a number of studies have contributed evidence on the positive effects this feedback type may have on written production, not only when concerning peer feedback (e.g., Tuzi, 2004) but also in terms of teacher-led electronic feedback (Ducate & Arnold, 2012; Elola & Oskoz, 2017). In addition, research focusing on the comparison of both asynchronous and synchronous teacher-led electronic feedback, albeit scarce, has again provided positive results in terms of language acquisition, with the majority of studies again mainly concerning peer corrective feedback (Honeycut, 2001; Liu & Sadler, 2003; Ware, 2004).

Due to the limited research available, this doctoral dissertation aims to add to the field by investigating and comparing traditional, delayed pen-and-paper WCF with digital, teacher-led asynchronous WCF. To provide the necessary background, the following section provides an overview of the empirical research that has compared teacher-led, asynchronous digital WCF with pen-and-paper WCF.

- *Delayed Asynchronous Teacher-led Digital WCF versus Traditional Pen-and-Paper WCF*

Little empirical research has specifically focused on the provision of delayed asynchronous teacher-led WCF, particularly in comparison to more traditional pen-and-paper feedback methods. The studies available have shed light on the effects of providing digital types of

feedback in comparison to more traditional types of feedback (Ene & Upon, 2014; Tafazoli et al. 2014). What follows is a summary of the main findings.

Table 3. Summary of Research on Teacher-led Asynchronous Digital WCF versus Traditional CF

STUDY	PARTICIPANTS	DESIGN	TASK	TIME SPAN	FINDINGS
ENE & UPTON (2014)	12 undergraduate non-native English speakers	Teacher-led feedback using the Word review function	2 writing assignments (with their corresponding drafts): text summaries (Word document)	Several weeks (one composition course)	Focus of teacher-led feedback remains the same for both modalities. Electronic feedback proved beneficial for writing accuracy
TAFAZOLI ET. AL (2014)	86 Tourism undergraduates	Teacher-led Direct unfocused WCF (mainly on grammatical errors) provided digitally and/or on pen-and-paper	5 writing assignments (Word documents/ pen-and-paper)	10 sessions during the second semester	Participants improved more in an on-line feedback environment (more influential for participants and more positive effects on grammatical accuracy)

Ene and Upton (2014) explored the differences between teacher-led electronic feedback and handwritten feedback in an attempt to compare and contrast the focus teachers put on error corrections and to view how and to what extent the error corrections affect the overall accuracy of the written assignments. Therefore, the primary focus of the study was to look at how providing teacher-led feedback in a digital environment might differ from doing so in a pen-and-paper mode. Data was collected in a U.S. university programme and 12 non-native English speakers participated in the study. The participants belonged to an undergraduate degree in

engineering and were all enrolled on a EAP programme. The writing data collected consisted of two essays and their corresponding drafts. The essays and their drafts were handed in electronically and the teacher provided feedback by using the review function in Word. Students writing was coded according to their uptake —successful, unsuccessful, unattempted, and unverifiable (p. 85)— (by comparing writing 1 and 2). Results from this study were compared to the existing empirical research on traditional WCF including Ferris et al. (1997) and Sommers (1982). Findings suggest that teacher-led digital WCF varied very little in terms of the focus when compared to traditional handwritten WCF. In addition, it was found that electronic feedback elicited revisions not only on grammatical errors but also on errors related to content and organization (p. 86). Therefore, the implementation of digital WCF within a classroom was proven to be an effective method of feedback provision.

In the same year, Tafazoli et al. (2014) published a study exploring the effects of electronic feedback on the grammatical accuracy of ESP writing when compared to traditional paper feedback. In addition, the study also focused on the perceptions and attitudes of students towards electronic feedback practices. Data was collected from 86 ESP Iranian students enrolled on a Tourism degree at the University of Applied Science & Technology (Mashhad). Participants were divided into two groups —one class writing online and another writing on paper—and were asked to submit 5 writing assignments on which they received grammatical feedback (p. 357). Results from this study showed that not only did students prefer electronic feedback, those who received it also performed better in terms of their grammatical accuracy when compared to the print experimental group, proving again, as in Ene & Upton (2014) that digital feedback is beneficial for L2 learning.

As shown in Table 3 and summarised above, comparative research on digital and pen-and-paper teacher-led WCF is scarce. Yet results points to benefits for the inclusion of digital feedback types in the L2 classroom. However, to the best of our knowledge, no research is available on the processing stage of WCF in the two writing environments (digital and pen-and-paper) and therefore, there is a need to address the ways in which these two writing conditions may affect the processing stage of feedback provision. The multimodal nature of digital feedback, which allows for feedback to be provided via text, audio or even video (Elola & Oskoz, 2016), opens up new horizons for writing processes and WCF processing research. As stated by Coyle, Nicolás-Conesa and Cerezo (forthcoming/2023), “research which could elucidate this underexplored domain is imperative to advance an underexplored yet growing area of interest”.

The next section will provide an overview of the research carried out thus far on WCF processing by exploring the methodologies implemented, the coding schemes utilised, and the results obtained in the existing empirical studies.

II.3.3.3. Written Corrective Feedback Processing

SLA-oriented WCF research has gradually shifted its orientation to studies that zoom in on the actual processing stage of feedback provision in an attempt to explore ways in which students process the feedback they receive and how, if possible, instructors can enhance learners' engagement with the WCF.

This growing research trend is working towards identifying the possible benefits metacognitive activities may have on a student's WCF processing as well as their consequent written products (Bitchener & Storch, 2016; Caras, 2019; Cerezo et al., 2019; Wigglesworth & Storch, 2012) and to confirm what was stated by Polio (2012) and later supported by Bitchener (2019) and Leow (2015) regarding WCF only being truly useful if students do something with it. The available research has shown that when a student/participant is presented with a metacognitive activity during the feedback processing stage, they will essentially perform better in the re-writing/second writing stage, and therefore advocates for this type of activity in the L2 classroom. The following section provides an overview of the research methods that have been used in L2 and WCF processing studies before zooming in on the empirical research relevant to the study, that is, research that has explored oral introspective measures and research that has explored written introspective measures.

- *Overview of Research on WCF Processing*

In general, research on L2 processing has mainly focused its attention on oral introspective measures and only a small number of recent studies have begun to address written methods as a means to explore L2 learning processes. Within this research, oral (think-aloud protocols, stimulated recalls, etc.) and written introspective measures (including written journals, written self-explanations, diaries, etc.) have been implemented in order to reveal how learners process L2 forms and to what extent they are aware of the input they receive. Studies on L2 processing can be divided into two main strands. On the one hand, research that focuses on language learning (e.g., Ishikawa & Révész, 2020; Ishikawa & Suzuki, 2016; Suzuki &

Brooks, 2009; Suzuki & Itagaki, 2007, 2009) and on the other, research that focuses on written corrective feedback processing (e.g., Caras, 2019; Cerezo et al., 2019; Kim & Bowles, 2019; Manchón et al., 2020; Qi & Lapkin, 2001; Sachs & Polio, 2007; Suzuki, 2012, 2017; Swain & Lapkin, 2002; Wigglesworth & Storch, 2012a,b).

Due to the focus of this doctoral thesis on written corrective feedback provision, an overview of the research available on written corrective feedback processing is provided below. This available research has explored a range of methodological procedures including the use of think-aloud protocols and stimulated recalls (verbally mediated data) as well as the implementation of noticing tables, diaries and written languaging tables (written mediated data) in order to respond to three main aims:

(i) Establish levels of feedback processing (depth of processing)
(ii) Investigate engagement with the WCF and subsequent revisions in re-writings and post-tests
(iii) Explore the effects of the levels of feedback processing on the accuracy of immediate text revisions

Before analysing the empirical research carried out on WCF processing and the results obtained, it is important to first explore some of the procedures implemented when collecting introspective data. As previously mentioned, the research can be divided primarily between studies that collected written verbalisations and those that collected oral verbalisations.

The research that has been carried out on WCF processing via written introspective measures can be divided according to two different types. On the one hand, studies that implemented non-concurrent procedures, which involve learners providing retrospective verbalisations on the task they have completed. As an example, the participants in Ishikawa & Révész, (2020) were asked to compare their writing to a model text and later discuss, in writing, their thinking process. On the other hand, there are studies that have used concurrent methods which involve learners providing written verbalisations whilst completing another task. As an example, Nicolás-Conesa et al., (2019) asked their participants to complete a written languaging table whilst processing the WCF provided on their writing. In these studies, learners are asked to discuss the error corrections they have received, and in some cases, provide metalinguistic explanations for those errors (e.g., Manchón, et al., 2020). Although written verbalisations have been used to shed light on L2 learner processes, some authors have questioned their veridicality for collecting data on the cognitive processes of L2 learners as

they “may not be closely related to underlying thought processes” (Ericsson & Simon, 1993, p. 109). In addition, and regarding the use of written languaging, some empirical research has concluded that despite a potential tool for language learning (Suzuki et al., 2023), written languaging tables “essentially address the product of WCF processing and not the process itself, so the insights they offer into levels of awareness or DoP should be considered with this proviso in mind” (Coyle et al., 2023). As previously mentioned, the present thesis includes a central methodological aim of exploring the affordances of written languaging and think-aloud protocols as WCF processing instruments capable of uncovering the cognitive processes involved in WCF processing.

For the studies involving oral verbalisations of L2 learners whilst processing WCF, research has implemented the use of stimulated recalls, oral languaging and think-aloud protocols.

Stimulated recalls are a tool frequently used in SLA research in which learners are encouraged to verbalise what was going through their mind whilst completing a specific task. That is, this particular type of introspective measure is implemented *a posteriori*, once the participant has completed the task at hand and is therefore classed as a non-concurrent procedure. Therefore, upon finalising the experiment, the subject is presented with stimuli which may include a recording (aural/visual) of them completing the task or perhaps, written stimuli, including feedback, for example. Upon receiving the stimuli, participants are then asked to recall what they were thinking at the time of completing the task and the instructor may guide the recall session with specific questions or prompts. Most empirical research in the field of corrective feedback, implement stimulated recall as a means to gather information regarding student (or teachers) perceptions and/or their engagement with CF and therefore, many studies have used this non-concurrent procedure in conjunction with oral CF (e.g., Mackey et al., 2000; Nabei & Swain, 2002; Salazar, 2012), computer-mediated WCF (e.g., Koltovskaia, 2020; Shintani, 2015) and pen-and-paper WCF (e.g., Lira-Gonzales et al., 2021).

A second type of oral data collection procedure includes oral languaging. The term “languaging” was first introduced by Swain (2006) who defined this concept as “the process of making meaning and shaping knowledge and experience through language” (p. 98). As previously mentioned, sociocultural theory views language learning as a process in which verbalisations and *languaging* play a crucial role. Therefore, encouraging students to engage in languaging prompts learners to work through any L2 problems they may encounter and eventually find solutions to these issues (Suzuki & Storch, 2020). Two types of languaging have been distinguished by Swain, including: (i) other-directed talk, also known as

collaborative dialogue, that is, talking to others, and (ii) self-directed talk which is often referred to as private speech, which would involve a learner talking to oneself (Swain & Watanabe, 2013). As opposed to written languaging, which as previously mentioned, has been found to be a reflection of the product of the WCF processing task, rather than a representation of the cognitive processes, oral languaging may be able to provide more valuable data on learner's attentional processes, levels of awareness, depth of processing and use of cognitive strategies, as the reactive elements written languaging may present (i.e., learners having to divide their attention between two tasks simultaneously) are eliminated for oral languaging, and learners are able to verbalise their thoughts whilst reflecting on the WCF provided, without being potentially distracted by a written languaging table, for example.

A third type of elicited oral verbalisations include think-aloud protocols. This introspective measure generally encourages L2 learners to verbalise what is going through their mind whilst they are completing a specific task or a very short time after task completion. Thus, think-aloud protocols can be divided into two main types which include: (i) concurrent think alouds and (ii) retrospective think alouds. This division was established by Ericsson and Simon (1993) in the field of cognitive psychology and refers to the distinction between when a think-aloud is completed simultaneously to the task at hand (concurrent), and when a think-aloud is requested upon finalising a task (retrospective), the main difference being "the time-lag that is present in the latter, which is often regarded as having influence on the accuracy and completeness of the data" (Zhang & Zhang, 2020, p. 303). However, the effectiveness of concurrent verbalisations in eliciting introspective data have been confirmed by researchers including Fox et al. (2011) who confirmed that this type of think-aloud does not alter task performance and is therefore, a true reflection of the cognitive processes involved in task completion.

In addition to the distinction between concurrent and retrospective verbalisations, Ericsson and Simon (1993) also categorise think alouds according to the information they elicit. Accordingly, they distinguished between three types of verbalisations, or levels, as the authors refer to them, which differ according to whether or not the verbalisation consists in simply thinking aloud without any additional requirements or specifically asking the participant to provide precise information, such as explanations, or justifications, etc. The first type is referred to by Bowles and Leow (2005) as non-metacognitive verbalisations, whereas the latter level of verbalisation (level 3) corresponds to what the same researchers have defined as metalinguistic think-alouds (Bowles, 2010; Bowles & Leow, 2005), or metacognitive verbal reports (Bowles, 2010), which can be directly comparable to what is expected in written languaging tables.

Despite the ever-growing use of think-alouds in SLA research, the question of its validity as a true representation of cognitive processes has been a constant point of interest. Thus, many studies have investigated the veridicality of this measure and the possibility it has of causing reactivity in empirical research on L2 processes. Specifically, in the field of L2 writing, a number of empirical studies have explored the issue of reactivity in relation to L2 writing processes (Yang et al., 2014 and Yang et al., 2020), as well as in conjunction with L2 corrective feedback (Adrada-Rafael & Filgueras-Gómez, 2019; Sachs & Polio, 2007; Suh, 2020). Concerning L2 writing processes, Yang et al., (2014, 2020) found think-aloud protocols to have a negative effect on specific elements of L2 written production, including fluency and aspects of syntactic and lexical complexity. However, no reactivity was found in relation to overall quality of writing, including L2 writing accuracy and text elements such as content and organization. Regarding the research that explored reactivity and L2 WCF (Adrada-Rafael & Filgueras-Gómez, 2019; Sachs & Polio, 2007; Suh, 2020), results were mixed. No reactivity was found when using concurrent verbalizations for processing WCF in the studies by Adrada-Rafael and Filgueras-Gómez (2019) and Suh (2020). Yet, in the 2007 research by Sachs and Polio, TAs were found to be reactive in relation to immediate post-test revision accuracy. Differences in the variables included in the study, including the language (and proficiency) used when processing via TA's could be the answer to such variances in results and inconsistent findings in reactivity research in general. As Bowles (2010) concluded in her meta-analysis on reactivity studies in the field of psychology and SLA, conclusions are extremely difficult to make due to the dependency of results on innumerous variables that are at play in this field of research.

Given that most research has found positive benefits for the implementation of TAs and other (non)concurrent oral verbalisations, a range of studies exploring oral introspective measures as a written corrective feedback processing instrument (including stimulated recalls, collaborative dialogues, and TAs) are available.

In the following section, an overview will be provided of the research that implemented oral verbalisations as a WCF processing measure and research that included written verbalisations as a means to explore WCF processing. These overviews will look into the methodologies implemented in the studies, including the types of feedback researched, the coding schemes used for the introspective measures and overall results.

- *Empirical Research on Oral Verbalisations and Written Corrective Feedback Processing*

One of the earliest written corrective feedback studies in which oral verbalisations were used was a case study by Qi and Lapkin (2001), in which they introduced think-aloud protocols as an introspective measure to tap into the cognitive processes and the amount of noticing their two participants demonstrated when comparing their written texts to reformulated versions. The two native Mandarin ESL learners were asked to complete a three-stage writing task in which they responded to open-ended picture prompts and wrote story compositions. They received reformulations as the feedback method and compared them to their original texts whilst thinking out loud. Qi and Lapkin (2001) coded the TAs by looking at language related episodes (LREs) defined in this particular study as "a segment […] in which a learner noticed a language-related problem he/she encountered while comparing his/her text to a reformulation and addressed it either by accepting the reformulation and providing a reason, or only noticing the difference without giving a reason" (p. 287). In addition, any instances in which self-corrections were made, but not verbalised in the TAs, were also considered, in this case, as language-related noticing. The LRE's were then coded according to three categories: (i) lexical, (2) form, and (3) discourse. The quality of the LRE's in terms of noticing was classified as either (i) *perfunctory* which refers to noticing without providing any reasons and (ii) s*ubstantive* which refers to noticing with reasons provided. Results from this case study show that "language related noticing may contribute to the improvement of L2 writing" (p. 294), with most LREs leading to subsequent improvements in the final written text. These positive results were interpreted as suggesting that language-related noticing, whether in the initial production stage, or the feedback processing/comparison stage, may lead to the possibility of retrieval in later stages of a writing task (p. 294). Additionally, the authors stated that the quality of noticing (with/without reasons) was a crucial variable and may vary as a function of the feedback type provided (reformulations in this particular case). Therefore, noticing can be enhanced so as to ensure learners notice the gap between their original texts and the corrected versions. Finally, concerning L2 proficiency, Qi and Lapkin (2001) speculated that learners with a higher L2 proficiency may find it easier to notice gaps in their L2 knowledge and consequently, be able to verbalise and problem-solve which evidently leads to higher benefits for L2 writing development.

As in the previously mentioned study by Qi and Lapkin (2001), Swain and Lapkin (2002) also looked into reformulations as a feedback instrument but this time exploring the

effects of collaborative dialogue as an oral languaging instrument. Two participants took part in a pre-test/post-test study in which, as a pair, they completed a jigsaw task, first orally and then in written form. The pair were then given a reformulated version of their original text and were asked to highlight the differences they noticed between this text and their original piece of writing, producing collaborative dialogue whilst doing so. A stimulated recall session was also included in which the participants were shown a video of their noticing stage and were asked to comment on certain verbalisations. Finally, for the post-test, the pair were asked to re-write their original text, this time individually. The verbalisations from the collaborative dialogue were coded for LREs, defined in this study as, "any part of the dialogue where learners talk about the language they are producing, question their language use, or correct themselves or others" (p. 292). As in Qi and Lapkin (2001), the LREs were then coded according to the same three categories: (i) lexical (ii) form, and (iii) discourse. The transcriptions from the stimulated recalls were also coded for LREs, specifically looking at whether the students had accepted or rejected the reformulations noticed. Results from the study showed that the introduction of reformulations enhanced the noticing stage of the study and were effective in engaging the pair in collaborative dialogue in which they discussed and reflected on the gaps between their original texts and the reformulated version. In addition to the high engagement with the corrections, the opportunity to language orally also played a beneficial role on the subsequent text revision as around 78% of the error corrections were incorporated into the revised texts.

Adding to the research on reformulations in conjunction with oral verbalisations, Sachs and Polio (2007) explored the cognitive processing of 15 English learners whilst processing two different types of feedback following a repeated-measures design. One group received error corrections on their writing, a second group received reformulations and a third group received reformulations but were asked to process them via think-aloud protocols. The think-aloud protocols were coded according to the noticing and two categories were established based on the amount of explanation provided. Therefore, they distinguished between, (i) noticing with metalanguage/reasons and (ii) noticing (without any further reasons provided). In terms of the results, surprisingly, they found that the silent processing group outperformed the think-aloud protocol group, therefore suggesting that the implementation of think-aloud protocols can have a negative reactive effect on feedback processing. Hence, they suggested that despite being a valuable instrument for collecting data on the cognitive processes involved in processing feedback, they should be used with caution as an instrument to enhance language learning via feedback processing. In addition, they found that the group who received error

corrections, outperformed the groups who received reformulations which was unexpected when compared to Qi and Lapkin's (2001) positive findings for the use of reformulations. On a more positive note however, they were able to confirm that any verbalisation of an error, would likely lead to a change and improvements in text revision.

Due to these varied results, Sachs and Polio (2007) decided to carry out a second experiment, including the same experimental feedback groups as before, but also introducing a control group. In addition, the repeated-measures aspect of the first experimental study was eliminated and more time was left between the feedback processing and revision stage. In this second study, which included a higher number of participants (54 ESL university students), results showed that again, the error correction group were the ones to produce the most accurate revisions, with the control group producing the least. However, unlike in the first experiment, the second study showed no significant differences between the reformulation groups (+/- think-aloud protocols), proving once again that the provision of feedback over no feedback is beneficial for L2 learners. In terms of noticing, the second study also confirmed the benefits of think-aloud protocols, confirming their facilitative qualities when it comes to noticing gaps in linguistic knowledge. Furthermore, instances in which noticing with reasons occurred in the feedback processing stage, were consequently associated with successful changes in subsequent rewritings.

A few years later, Wigglesworth and Storch (2012a) expanded on the research carried out on oral verbalisations and written corrective feedback, by exploring the effects of indirect feedback, in the form of editing symbols, as well as reformulations (as seen in Qi & Lapkin, 2001; Sachs & Polio, 2007; Swain & Lapkin, 2002). Additionally, rather than looking at individual writing, they explored the effects of collaborative writing and collaborative dialogue as an introspective measure (as seen in Swain & Lapkin, 2002). 72 participants were involved in the study in which, in pairs, they were invited to write a report on which they received written corrective feedback five days later. Participants were asked to discuss the feedback provided by engaging in collaborative dialogue with their partners. The pair talk was analysed according to LRE's, (as done in Qi & Lakin; Swain & Lapkin, 2002) and were categorised according to the following three types: (i) Form (ii) Lexical and (iii) Mechanical. In addition, noticing was categorised according to the interaction between the pairs, therefore, in instances in which two or more turns were involved between the pair, the interaction was classified as engaging. In change, no engagement was considered if the LREs consisted in just one turn, with no interaction between the pair. If over 70% of the total LRE's consisted in more than one turn, this was classified as high engagement, for 40-69%, medium and, for anything below 40%,

low engagement (p. 86). Results show that both of the feedback types implemented led to improvements in accuracy, with the reformulation group outperforming the indirect WCF group. In relation to the level of engagement with the feedback, the editing indirect group outperformed the reformulation group when it came to accepting and engaging with the feedback provided. However, no clear relationship was found between these levels of engagement and the subsequent incorporation of the error corrections into revised texts (which were completed individually, four weeks after the processing stage), with the more direct, reformulation feedback group demonstrating more enduring language learning, as manifested in subsequent new writing. In terms of the collaborative writing aspect, Wigglesworth and Storch (2012a) postulated that encouraging students to work in pairs and engage with written corrective feedback collaboratively may lead to greater language learning gains and writing development.

Caras (2019) shed further light on oral verbalisations and WCF research by carrying out a study in which 40 participants were invited to complete a descriptive composition online. This study included three experimental feedback conditions in order to explore to what extent different feedback types can affect the processing stage. Therefore, there was a direct WCF group, an indirect WCF group (as seen in Wigglesworth & Storch, 2012a), a metalinguistic WCF group and a control group and all participants were asked to think-aloud whilst processing their feedback. In order to measure the processing participants had undergone whilst viewing their feedback, Caras (2019) coded the languaging data according to the levels of depth of processing (DoP) and categorised each language related episode as either high, medium, or low DoP. Results demonstrated that the direct WCF promoted the least amount of deep processing, with the indirect group only processing errors at a low level, whereas the metalinguistic feedback group, were the ones who processed their errors deeply. Despite this, the direct group was the experimental condition who performed best on the revised writing (draft 2) in terms of accuracy, especially when it concerned *ser* versus *estar,* with the metalinguistic group also outperforming the control group. However, it was found that in the delayed post-test, none of the unfocused feedback groups played a lasting role on learner's L2 accuracy.

In the same year, Kim and Bowles (2019) carried out a study with 22 participants in which they explored the effects of two types of written corrective feedback, including reformulations, as in previous research (Qi & Lapkin, 2001; Sachs & Polio, 2007; Swain & Lapkin, 2002; Wigglesworth & Storch, 2012) and direct written corrective feedback, as seen in Caras (2019). Participants were asked to complete two argumentative essays on which they

received both types of feedback mentioned in a counter-balanced manner. That is, if on one essay they received direct WCF, on the second essay, they then received reformulations and vice versa. Upon receiving the corresponding feedback, participants were asked to think-aloud and process the error corrections provided on their essays. Processing instances were coded according to whether they were high or low in terms of the levels of depth of processing. Regarding the results of the study, authors found that reformulations prompted participants to process errors at a deeper level than in the direct WCF group. In turn however, and unlike in Caras (2019), it was the direct WCF processing group who noticed more errors and commented on them more frequently. Therefore, a trade-off effect was found in which, despite the reformulation group leading to less noticed errors, the errors that were noticed were processed at a deeper level. In terms of L2 accuracy, results in this study were not measured according to L2 learning gains.

Finally, and adding further research to the field, Adrada-Rafael and Filgueras-Gómez (2019) also looked at the processing of reformulations but they included the language of the think-aloud protocols the participants did as a variable. That is, they had an experimental group that was requested to think-aloud in their L1 (English) and an experimental group that were asked to think-aloud in their L2 (Spanish), in order to view whether the language used whilst processing reformulations, played a role on the depth of processing of the feedback. 29 advanced-level Spanish learners were involved in the study and were asked to complete a re-telling of a picture story (the same task implemented in Sachs & Polio, 2007). One group processed the reformulations received on this task in their L1 and the second group processed their errors in their L2. In terms of the results found, unlike in Sachs and Polio (2007), no reactivity was found for both experimental groups, suggesting that the implementation of think-aloud protocols with advanced L2 learners should be encouraged. Additionally, they found that more instances of processing were made in the L2 languaging group. However, in terms of the depth of processing of the language instances, a deeper level was found in general, amongst the L1 think-aloud condition. More specifically, learners who processed their errors in the L2, produced more grammatical and lexical instances of processing, when compared to the students who processed in their L1, but authors remind readers that the participants were advanced level Spanish students and thus, these somewhat surprising results are most likely due to their high level of linguistic proficiency.

It is clear from this overview of research that although it appears that L2 WCF processing can lead to improvements in subsequent writing, many variables are at play when attempting to measure the processing of corrective feedback and its benefits. First, there's a

significant variability in the introspective measures used and the coding scheme that each study has implemented, with some studies focusing more on noticing (Qi & Lapkin, 2001; Sachs & Polio, 2007), others on engagement and LRE types (Swain & Lapkin, 2002; Wigglesworth & Storch, 2012a) and others on levels of depth of processing (Adrada-Rafael & Filgueras-Gómez, 2019; Caras, 2019; Kim & Bowles, 2019). In addition, within these different coding schemes for the data, there are differences in the levels established by each author, therefore in the research that looked into levels of DoP for example, some authors have included three levels (Adrada-Rafael & Filgueras-Gómez, 2019; Caras, 2019), whereas others have included two levels of DoP (Kim & Bowles, 2019), and in the research that focused on noticing, discrepancies can be found in the coding of the noticing instances, with some authors categorising them according to noticing with/without reasons (Sachs & Polio, 2007, experiment a and b) and others, according to perfunctory or elaborate noticing (Qi & Lapkin, 2001). These differences in introspective measures and coding schemes used, make it more difficult to compare the results and attribute the benefits to specific types of oral verbalisations. In addition, the differences in coding schemes used reflects the difficulty of categorising the cognitive processes of learners whilst processing their WCF (see McBride & Manchón, 2023 for a detailed reflection on the process of coding WCF processing data), and "agreement as to the precise nature of DoP would facilitate more rigorous methodological decision-making, including the selection or combination of appropriate data collection instruments (Coyle et al., 2023).

In terms of the populations included in the research, most studies have been carried out with higher-education students (Adrada-Rafael & Filgueras-Gómez; Caras, 2019; Kim & Bowles, 2019; Qi & Lapkin; Sachs & Polio, 2007; Wigglesworth & Storch, 2012), with only one study focusing on high-school students, a more under-represented population (Swain & Lapkin, 2002). This is in line with most research on feedback processing in which most participants come from an undergraduate or postgraduate setting. Due to this, the L2 proficiency level of most participants in the body of research reviewed, corresponds to intermediate to advanced levels. As the most common population under study comes from a higher-education background, this also generally leads to the participants coming from a background in language and linguistics, something which scholars have called to be addressed in future research (Coyle et al., 2023; Manchón & Leow, 2020)

Interestingly, only three studies included a control group (Caras, 2019; Sachs & Polio; 2007, experiment b; Wigglesworth & Storch; 2012), with the study by Caras (2019) including the control group in terms of the feedback provided, and not related to the think-aloud protocols

(all groups were asked to think-aloud). Thus, any differences found in the research available, attributable to the oral verbalisations must be considered tentatively, as without a group of participants who complete the tasks silently (Leow & Bowles, 2023), it becomes extremely difficult to know whether or not the introspective measures are the tools responsible for the results obtained.

Relevant to the present doctoral thesis is the writing environment in which the studies took place. As previously mentioned, WCF processing research in a digital setting is scarce and in the empirical research reviewed above, only two studies (two of the most recent) took place in a digital setting (Adrada-Rafael & Filgueras-Gómez, 2019; Caras, 2019). Interestingly, the study by Adrada-Rafael and Filgueras-Gómez (2019) implemented the same writing task as the one used in Sachs and Polio (2007), which consisted in a picture-based prompt in which participants had to recount the story depicted in the six pictures, the only difference been that participants in Sachs and Polio (2007) completed the task on pen-and-paper and the participants in Adrada-Rafael and Filgueras-Gómez (2019) completed the task on a computer, in Microsoft Word. The feedback of choice for both studies was also the same, reformulations, yet the results found in terms of the introspective measure used were different. Sachs and Polio (2007) found the think-aloud protocols to be reactive, as the experimental group who processed the reformulations without thinking aloud, performed better. In change, Adrada-Rafael and Filgueras-Gómez (2019) didn't find the think-aloud protocols to be reactive in their study. One of the main differences between the two studies is the writing and L2 processing environment in which they took place and therefore, it could be one plausible explanation for the differing results and grants further investigation.

Despite the discrepancies in methodological approaches in the body of studies reviewed, research has been able to confirm the benefits of oral verbalisations during WCF processing not only on L2 learners' noticing but also on their L2 writing accuracy.

Table 4. Summary of Research on Oral Instruments for Written Corrective Feedback Processing

STUDY	CONTEXT AND PARTICIPANTS	WRITING TASK	FEEDBACK TYPE	FEEDBACK PROCESSING INSTRUMENT	CODING SCHEME FOR PROCESSING	RESULTS
QI & LAPKIN, 2001	Two Mandarin ESL learners	Three-stage Writing task: Open-ended picture prompts	Reformulations	Think-aloud protocols	Noticing: (1) Perfunctory (2) Elaborate	Reformulations lead to language related noticing and L2 writing improvement
SWAIN & LAPKIN, 2002	Two grade 7 French immersion students (Canadian middle-school)	Collaborative information-gap picture-based jigsaw Task (narration) first orally, then in writing	Reformulations	Collaborative dialogue & stimulated recall	LRE's: (1) Lexical (2) Form (3) Discourse And: (1) Acceptance or (2) Rejection of the Error Corrections	Reformulations enhanced noticing and led to LRE's
SACHS & POLIO, 2007 (experiment 1)	15 high-intermediate English learners (university level)	30-min picture-description task	(1) Error corrections (2) Reformulations (3) Reformulations and Think-aloud protocols	Think-aloud protocols (L2)	(1) Noticing (2) Noticing with reasons and/or metalanguage	(1) Think-Alouds were reactive (Reformulation group without TAs included more revisions than the reformulation + TA group (2) Error correction group outperformed reformulation group (3) Noticing with metalanguage/reasons led to more revisions

SACHS & POLIO, 2007 (experiment 2)	54 ESL students (university level)	30-min picture description task	(1) Error corrections (2) Reformulations (3) Reformulations and Think-aloud protocols (4) Control group	Think-aloud protocols (L2)	1) Noticing (2) Noticing with reasons and/or metalanguage	(1) All groups outperformed the control group (2) Error correction group performed best in overall text revision (3) Noticing in the TAs were directly related to revisions
WIGGLESWORTH & STORCH, 2012a	36 EFL undergraduate pairs (advanced level)	Collaborative writing task: Report (graphic prompt)	(1) Reformulations (2) Indirect Feedback (3) Control group	Collaborative dialogue	LRE's: (1) Form (2) lexical (3) Mechanical And engagement: (1) High (2) Medium (3) Low	(1) Indirect feedback group outperformed the reformulation group in terms of engagement (2) No clear relationship between engagement and task revision. (3) Both feedback types led to improved accuracy, reformulations more effective short-term.
CARAS, 2019	40 Spanish learners (University)	Descriptive composition task (blog post) on a computer	Unfocused (1) Direct WCF (2) Indirect WCF (3) Metalinguistic WCF (4) Control group	Think-aloud protocols	Levels of DoP: (1) High (2) Medium (3) Low	(1) Direct WCF led to least amount of deep processing (2) Indirect WCF led to low levels of processing

						(3) Metalinguistic group led to deep levels of processing (4) Direct WCF group performed best on rewriting (accuracy) (5) No long-terms benefits (delayed post-test) for accuracy found for any unfocused feedback group.
KIM & BOWLES, 2019	22 high-intermediate undergraduates (enrolled on an ESL academic writing course)	Two Argumentative essays	(1) Reformulations (2) Direct WCF (Counter-balanced)	Think-Aloud Protocols	Levels of DoP: (1) High (2) Low	(1) Reformulations prompted deeper processing levels (2) Direct WCF provoked more noticing
ADRADA-RAFAEL & FILGUERAS-GÓMEZ, 2019	29 Advanced level Spanish undergraduates	Re-telling of a picture story (Sachs & Polio, 2007) using Microsoft Word	Reformulations	Think-Aloud Protocols: (1) L1 (English) (2) L2 (Spanish)	Levels of DoP: (1) Deep (2) Intermediate (3) Low	(1) More number of languaging episodes in the L2 languaging group (2) Deeper processing in the L1 languaging group

- *Empirical Research on Written Verbalisations and Written Corrective Feedback Processing*

In terms of the studies that focus their attention on written verbalisations and written corrective feedback, despite attracting less attention when compared to oral techniques, new research has recently been added to the field including Cerezo et al. (2019), Manchón et al. (2020), Moradian et al. (2020) and Simard and Zuniga (2020). The more recent publications are included in a volume edited by Suzuki and Storch (2020) on *languaging in language learning and teaching*.

In one of the first studies in this research domain, Suzuki (2012) explored the effects of written languaging in conjunction with written corrective feedback by inviting 24 Japanese intermediate learners of English to reflect on their error corrections via a written languaging sheet. Participants completed a TOEFL writing prompt and received direct written corrective feedback on all errors. They were asked to reflect on the corrective feedback and were encouraged to write down on a separate sheet of paper, why they believed their linguistic forms had been corrected. The languaging activity was done in the participants L1 (Japanese). Languaging data was coded according to written language episodes (WLE), which were defined in this study as languaging episodes "about each linguistic error that had been overtly corrected by the native English instructor". (p. 10) Each WLE was then narrowed down according to the three following groups (as in the oral languaging studies by Swain & Lapkin, 2002; Wigglesworth & Storch, 2021a): (i) lexis, (2) grammar and (iii) don't know, with any WLEs which didn't fit within these three broad categories, coded as (iv) other. Results from this study showed that learners tended to language most about grammatical errors rather than lexical ones. In addition, authors found that WLEs were very likely to be incorporated into subsequent text revisions (as found in previous research on LREs, Qi and Lapkin, 2001; Swain and Lapkin, 2002), and therefore lead to accuracy improvement both for lexical and grammatical errors. A few years later, in 2017, Suzuki set out to explore the direct effects of the quality of written languaging on L2 writing improvements. To this end, he coded the WLE's (written in the student's L1) into three categories (p. 12) including: (i) noticing only (i.e., not providing explanations (equating to Qi and Lapkin's, 2001 *perfunctory noticing*), (ii) noticing with reasons (showing an understanding of the error corrections, equating to Qi and Lapkin's, 2001 *substantive noticing*) and (iii) uncertainty (instances in which students were not sure of why the error correction has been provided). Results from this study showed that all instances of noticing (both with or without reasons) led to improvements in L2 writing accuracy. In

addition, following the provision of direct WCF, students were more likely to provide reasons for the error corrections, representing a deep level of awareness.

In the same year, Moradian, et al. (2017) added further evidence in favour of written verbalisations by contributing to the body of research on direct WCF processing via notetaking. Their study in which 38 Iranian low-intermediate EFL adolescent learners took part, invited participants to discuss the errors identified in their dictogloss writing task. The processing task consisted in noting down (in their L1 or L2) any reasons for the errors either on a separate piece of paper, or below the original writing itself. The aim of this study was to explore the effects in terms of accuracy in revised texts so no information regarding the coding of the notetaking was provided, as this was not included in the analysis of the empirical study. Results corroborate with previous research (Suzuki, 2012) as authors found that participants who received direct WCF and processed it in written form, outperformed the group who received just direct WCF (without notetaking). However, both treatment conditions were able to enhance participants L2 grammatical accuracy in revised texts, proving once more the effectiveness of providing direct WCF for L2 learners.

Cerezo et al. (2019) contributed to the field of written languaging by looking at the effects of two types of WCF: direct and indirect respectively. Forty-six participants took part in a pre-test/treatment/post-test study in which they performed a time-compressed argumentative writing task and received direct or indirect feedback in the experimental conditions, with a control group not receiving feedback on the writing task. Written languaging was measured using a table in which, following the instructions given in Suzuki (2012), participants were asked to provide the error transcription, the correction, the type of error (category) and a metalinguistic explanation for the error corrections provided. The written languaging in this study was coded according to five levels of depth of processing, depending on the amount of information provided in the written languaging tables, constituting a representation of the cognitive effort expended by students when processing WCF. These levels were then coded according to the levels of awareness of students and three categories were established (following Leow, 1997) including, (i) Level 1- awareness at the level of noticing, (ii) Levels 2 and 3- awareness at the level of reporting and (iii) Levels 4 and 5- awareness at the level of understanding (p. 180). In terms of the results from the study, deeper levels of feedback processing were related to the direct written corrective feedback group. However, and as found in Suzuki (2017), the actual levels of processing (high, medium, or low) did not play a role on the accuracy of the revised texts, but they did however determine the choices participants made when incorporating or deleting their original errors into the

rewritings. Nevertheless, both feedback groups (direct and indirect) led to significantly more noticing than the control group, with the indirect group noticing more than the direct in all of the error categories. This raises the issue again, as in the research carried out on oral languaging, that although some feedback types may lead to more noticing (e.g., indirect), others provoke less noticing but a deeper processing (e.g., reformulations as seen in Kim & Bowles, 2019). Taking into consideration the feedback types, it seems that learners are able to process errors at a deeper level when they are provided with the correction (as is the case for the direct WCF groups). In turn, the participants in the indirect WCF conditions notice the errors more but are less likely to process them at a deep level, likely due to them not being able to correct the error and not having the correction provided.

More recently, a collection of empirical studies investigating written languaging were brought together in a publication on *languaging in language learning and teaching* by Suzuki & Storch (2020). Follows is a summary of the main research findings on written languaging and written corrective feedback included in this volume.

Manchón et al. (2020) added to the body of research by exploring written languaging and written corrective feedback in two writing conditions, which included individual and collaborative writing. The participants (118 in total) were invited to complete a problem-solving picture-based writing task either collaboratively in pairs, or individually. The participants in the experimental groups were then provided with direct written corrective feedback on their errors and all participants were asked to complete a written languaging table in which, following Cerezo et al. (2019), they had to include the errors, their corrections, as well as a metalinguistic explanation for any errors and their corrections. Results from this study confirmed those found in Suzuki (2012, 2017) in which processing direct written corrective feedback via written languaging led to a higher detection of errors and was therefore more beneficial than languaging without the aid of corrective feedback. Additionally, and contrary to their initial hypothesis which considered the positive effects of collaborative writing from previous research (Wigglesworth & Storch, 2012a, 2012b), there were no statistically significant differences found between the individual and collaborative writing group where levels of depth of processing were concerned, as WCF processing in both groups led to high levels, in an equal manner.

Another empirical study included in Suzuki and Storch (2020) was the research carried out by Moradian et al. (2020). As in previous research (Moradian et al., 2017; Cerezo et al., 2019), the authors compared the effects of written verbalisations after the provision of direct and indirect written corrective feedback on all grammatical errors. Their participants (15 in

each experimental condition) completed a five-week data collection in which they completed a pre-test and post-test writing task, as well as five writing tasks (taken from their textbook) on which they received feedback and processed their errors in written form. Results once again confirmed, as in previous research (e.g., Manchón et al., 2020; Moradian et al., 2017), that the very act of languaging errors led to accuracy gains regardless of the type of feedback provided. However, direct WCF was found to be more effective in engaging participants in a higher level of metalinguistic awareness of their errors (similar to findings found in Cerezo et al., 2019 and Suzuki, 2017 in which the provision of direct WCF tended to lead to deep levels of processing).

As with the research on oral WCF processing, the results for the studies on written languaging and written corrective feedback are once again varied. One of the principal reasons behind this variation in results is due to the implementation of a wide range of feedback types, populations under study, L2 proficiency level of students, as well as differences in the introspective measures used and the coding schemes created.

Once again, the majority of participants included in the research were adult L2 learners (Cerezo et al., 2019; Manchón et al., 2020; Suzuki, 2012, 2017), with only the study by Moradian et al. (2017) including adolescent L2 learners. However, unlike in the oral verbalisation research in which the proficiency level was generally high (Adrada-Rafael & Filgueras-Gómez; Kim & Bowles, 2019; Sachs & Polio, 2007; Wigglesworth & Storch, 2012a), the participants' L2 level of proficiency in the written verbalisation studies reviewed is generally lower (Cerezo et al., 2019; Manchón et al., 2020; Moradian et al., 2017, 2020).

Many differences can also be found in the way in which the written data was coded, with some studies analysing the written verbalisations in terms of written language episodes (Suzuki, 2012; Moradian et al., 2020) others in relation to the depth of processing levels (Cerezo et al., 2019), and others in terms of awareness and noticing (Cerezo et al., 2019; Suzuki, 2017). These differences in the analysis of the written output of the processing stage make it extremely difficult to make generalisations in terms of the results, as each study is measuring a different phenomenon. Thus, as suggested for research on oral verbalisations, it is essential to reach a consensus on the coding schemes used in this research area to be able to make generalisations and truly discuss the effects of including written verbalisations in WCF processing. One study by Manchón et al. (2020) made a crucial advancement in this area by providing a range of methodological observations upon analysing the data output the written languaging tables provided. The authors concluded that the written languaging data did not correspond specifically to the students' cognitive processes whilst viewing their WCF (and their levels of DoP) but instead provided data on the outcome of the WCF processing stage.

These reflections are in line with suggestions made by Suzuki et al. (2023) in which they claim that some written verbalisations may not be a true reflection of the cognitive processes that are involved in L2 feedback processing. However, they do confirm that written languaging can and should be implemented as part of the learning process as it clearly holds potential in promoting language learning.

Given the methodological concerns raised about written verbalisations, together with the benefits both oral and written verbalisations seem to have on L2 writing gains, future research would likely benefit from analysing both written and oral languaging data produced whilst students process WCF as written languaging alone can "only partly capture the DoP taking place" (Manchón et al., 2020). In a recent publication worthy of mention by Coyle and Roca de Larios (2020), collaborative dialogue as an oral introspective measure strategy was used alongside written notes as a WCF processing technique. This innovative study explored the WCF processing of 16 young learners belonging to EFL and CLIL classes upon receiving model texts on their collaborative writing task. Students were asked to underline and take-down notes whilst comparing their original text, which was a response to a six-frame picture story and engage in collaborative dialogue at the same time. The LRE's were categorised according to whether they referred to (i) lexis, (ii) form, or (iii) sentences. In addition, they operationalised noticing in terms of whether or not students noticed a gap between their original texts and the models provided and a number of "strategies" were identified following a data-driven process which corresponded to how the students handled the model texts. (p. 5). Results showed that in both classroom contexts (EFL and CLIL) students mainly focused on lexical errors when comparing their texts to the models provided. In addition, the authors found CLIL students to be more capable of finding solutions to their problems within the model text, which subsequently led to a greater uptake for their revised texts. The authors therefore concluded that previous learning experiences (in this case EFL or CLIL) may play a role on how students cognitively process WCF. To the best of our knowledge, this is the first research article to combine both oral and written feedback processing instruments in the same study. However, we are not aware of studies comparing the differential effects oral and written feedback processing may have on levels of depth of processing and L2 writing accuracy, and thus, the present doctoral thesis aims to contribute to the field by exploring both modes as WCF processing instruments.

Table 5. Summary of Research on Written Languaging and WCF Processing

STUDY	CONTEXT AND PARTICIPANTS	WRITING TASK	FEEDBACK TYPE	FEEDBACK PROCESSING INSTRUMENT	CODING SCHEME FOR PROCESSING	RESULTS
SUZUKI (2012)	24 adult native Japanese speakers on an English composition course	2 TOEFL writing prompts	Direct WCF for all linguistic errors	Open-ended written languaging sheet (L1)	WLE's (1) Lexis (2) Grammar (3) "Don't know" (4) Other	(1) Grammar WLE's were most common (2) WLE's led to improved accuracy in subsequent revisions (3) WL did not play a different role in terms of different language areas being processed and learnt
SUZUKI (2017)	24 adult native Japanese speakers on an English composition course	2 TOEFL writing prompts	Direct WCF on all linguistic errors	Open-ended written languaging sheet (L1)	Awareness in the WLE's (1) Noticing only (2) Noticing with reasons (3) Noticing with uncertainty	(1) Direct WCF led to deep levels of awareness (2) awareness led to successful incorporations in the text revisions.

MORADIAN ET AL., (2017)	38 low-intermediate adolescent EFL learners	Dictogloss task	(1) Direct WCF + WL (2) Direct WCF	Notetaking (reasons behind their errors) on a separate sheet or underneath their writing task. (L1 or L2)	No details provided	(1) Both feedback types led to enhanced grammatical accuracy (2) Direct WCF + WL outperformed Direct WCF only group in terms of written accuracy on text revisions
CEREZO ET AL., (2019)	46 1st-year English studies undergraduates (B1 level)	Personal experience essay (Raimes, 1978)	(1) Direct WCF (2) Indirect WCF (3) Self-Edit	Structured written languaging table (L1)	DoP levels (1) Error transcription (2) Error transcription + Error correction or Error category (3) Error transcription + error correction and error category (4) Error transcription + error correction	(1) Noticing is enhanced by WCF provision (2) Indirect WCF led to higher noticing of lexis and an advantage over direct WCF for total amount of errors noticed (80% for indirect vs. 55% for direct)

					and metalinguistic explanation or error transcription + error category and metalinguistic explanation (5) Error transcription + error correction + error category and metalinguistic explanation Levels of awareness (1) noticing (2) reporting (3) understanding	(3) Direct WCF promoted deeper DoP levels (4) All levels of DoP led to corrections in text revisions
MANCHÓN ET AL., (2020)	118 undergraduate EFL learners (B1 level)	Individual (IW) and Collaborative (CW) completion of the Complex version of "*Fire Chief*" task (Gilabert, 2007)	(1) Unfocused direct WCF (2) Control group	Structured written languaging table (L1/L2)	Levels of cognitive effort: (1) Reporting (2) & (3) Reporting with limited elaboration	(1) Both WCF groups (IW & CW) outperformed control group processing WCF (2) Both groups successfully

					(4) & (5) Reporting with extended elaboration	incorporated correct revisions in post-test. (3) Direct WCF leads to high DoP in both conditions (IW & CW)
MORADIAN ET AL., (2020)	30 Iranian EFL learners (Low-intermediate)	5 Writing tasks taken from EFL textbook (*Four Corners)* + Pre and Post-test Writing prompt	Unfocused: (1) Direct WCF (2) Indirect WCF	Note-taking (L1/L2)	WLE's (1) correctly explained (2) incorrectly explained (3) unexplained	(1) Both feedback types + WL led to enhanced L2 accuracy

II.4. SUMMARY OF EMPIRICAL INSIGHTS AND OPEN QUESTIONS

The review of the literature presented in the previous sections provides convincing evidence of the important role that writing and written corrective feedback provision, processing, and use may play in instructed L2 learning. Although research varies in terms of, *inter alia*, focus and feedback types (as reviewed in II.3.3.), studies clearly show the benefits WCF can have on immediate L2 writing gains and longer-term language learning. However, as L2 classrooms are part of an ever-growing digital world, calls for research to include more digital environments is essential for a fuller picture of the intricacies of learning gains via writing and feedback processing and use across writing environments. Research (as reviewed in II.3.3.2) has confirmed that not only do students receive digital writing positively, but also that this type of writing environment is beneficial for L2 writing. However, research comparing traditional pen-and-paper writing with digital conditions is scarce and, therefore, further investigation into the impact digital writing may have, as compared to pen-and-paper writing, is warranted. This would facilitate shedding light on optimal writing conditions for L2 learners. It also remains to be explored how the writing environment may affect WCF processing, as well as whether any attested benefits for digital WCF provision also apply to the processing stage.

Regarding the available research on WCF processing (carried out generally in pen-and-paper environments), despite differences in the introspective measures implemented in the available studies (as reviewed in section II.3.3.3.), it is clear that providing students with the opportunity to process their errors, either by verbalising their thoughts orally or in written form, can lead to improvements in subsequent L2 writing and enhance engagement with the feedback provided. However, given the different insights each type of verbalisation provides, it is essential to expand on existing research by exploring the affordances (from both a methodological and a language learning point of view) of both oral and written introspective measures. As advocated by Manchón and Leow (2020) in their position piece on future research agendas, future empirical studies would benefit from more controlled studies in which the affordances of diverse instruments are tested, in order to fully capture the WCF processing stage. More recently, and adding to these calls, Coyle, Nicolas-Conesa and Cerezo (2023) have suggested expanding WCF processing research to include new modalities (e.g., digital environments) as well as enhance the comparability of findings by refining the constructs used and the coding schemes implemented.

In order to respond to these open questions and calls, the following part of the thesis (Part II) details the empirical study that was carried out. First the aims and research questions that guided the study will be stated, followed by a detailed explanation of the data collection and data analysis procedures implemented. The results are first presented and subsequently discussed according to the three research questions guiding the research.

PART II.

THE STUDY

CHAPTER III. AIMS AND RESEARCH QUESTIONS

The present PhD aimed to contribute to two relevant SLA-oriented lines of research reviewed in Chapter II: (i) effects of composing medium on written texts, on the one hand, and effects of feedback processing conditions on writing processes and products, on the other. As advanced in the Introduction, this intended empirical contribution includes a central methodological aim.

The motivation for these global aims derives from the considerations already discussed in greater length in Chapter II. On the one hand, in response to the mass shift seen in language classrooms to more online, digital learning environments, L2 writing scholars have advocated for research to explore and compare the effects of traditional pen-and-paper versus digital composing environments on the cognitive processes involved in writing and feedback processing, as well as on the characteristics of the resulting written texts (as more fully elaborated upon in II.2. in Chapter II). On the other hand, SLA-oriented research on written corrective feedback (WCF) has received ample attention throughout the years (as discussed in II.3, Chapter 2), with studies that focus specifically on the processing of WCF gaining increased attention more recently. We explained in II.3. that the construct of depth of processing (DoP) of WCF has become a key concern in theoretical and empirical feedback research. We also reported that studies in this domain have employed diverse methodological procedures -including think-aloud protocols and written languaging-, to obtain data on potential (i) effects of DoP on how deeply L2 users engage with the feedback provided on their writing; and (ii) correlations between DoP and language acquisition (operationalized in terms of improvements in text revisions). Yet, it was pointed out that scholarly debates have more recently focused on methodological considerations regarding data elicitation procedures in this research, with critics (e.g., Leow & Manchón, 2021; Manchón, 2023) calling for more controlled, methodologically-oriented studies in which the validity of the data collection instruments is tested, whilst also advocating for investigations in more diverse writing environments.

In response to these calls, the main aims of our research were to contribute empirically to previous research by exploring writing and feedback processing in both pen-and-paper and digital environments, and to shed light on the affordances of diverse introspective measures (individually and combined) commonly used for WCF processing. To achieve these global aims, the following research questions guided our study:

RQ.1 How does writing in a traditional pen-and-paper environment versus writing in a computer-mediated environment affect L2 written production in terms of CAF measures?

RQ. 2 How does the experimental manipulation during WCF processing affect L2 written production (in terms of CAF measures) in pen-and-paper versus computer-mediated writing environments?

RQ.3. How does the experimental manipulation during WCF processing affect L2 learners' levels of depth of processing of the feedback received in pen-and-paper versus computer-mediated writing environments?

CHAPTER IV. METHODOLOGY

IV.1. DESIGN OVERVIEW

The research conducted for this doctoral thesis followed a pre-test/treatment/post-test design. The independent variables were feedback processing conditions (there were 3 conditions according to the processing instruments used) and writing environment or composing medium (i.e., pen-and-paper vs. digital writing). The dependent variables were the levels of depth of processing (i.e., high, medium, low, and null levels) and the characteristics of the written texts (in terms of complexity, accuracy, and fluency indices). Participants were invited to write an initial text (pre-test) in time-constrained conditions. The writing task was the problem-solving, picture-based "Fire Chief" task (Gilabert, 2005, 2007), which was completed by half of the participants (18) online, via *GoogleDocs*, and by the remaining 18 participants on pen-and-paper. Regardless of the writing and processing conditions, all participants received unfocused, direct WCF on their initial written texts. The participants were then invited back to process the feedback received, according to the treatment group to which they were assigned: (i) think-aloud only, (ii) written languaging only, and (iii) simultaneous think-aloud and written languaging. The final task (post-test) invited participants back to rewrite their original text under the same conditions as in the pre-test.

IV.2. CONTEXT AND PARTICIPANTS

The study was carried out at the University of Murcia, Spain. The participants (n= 39; 30 females and 9 males) were undergraduate students in their fourth year of a degree in English Studies. Due to unforeseen circumstances, three participants had to be excluded from the study halfway through the data collection and, accordingly, the final number of participants included in the research was reduced to 36 (28 males and 8 males). Participants were all L1 Spanish speakers and their age ranged from 21 to 25 years (mean age 22). They were enrolled on an optional Applied Linguistics module as part of their undergraduate English Studies degree. They all volunteered to take part in the study outside of class hours and signed the corresponding informed consent. These participants were selected to take part in the study as they were the most accessible to the researcher at the time of the data collection. In addition, due to specific regulations at the university in which the data was collected, no monetary

rewards can be provided to potential participants and thus convenience sampling was the method adopted in order to ensure the participation of students. Participants did benefit from the established credit-bearing system for voluntary participation in empirical studies related to the modules in their degree studies. They were informed that participation was voluntary and that their decision not to participate would not have any negative effect on course grades. They were equally informed of the potential benefits to be obtained from having first-hand experience in the kind of empirical research that was so central in the teaching materials and course tasks in the Applied Linguistics module they were all enrolled in.

The participants completed an initial Oxford Placement Test (OPT) in order to confirm homogeneity in terms of proficiency level. Upon completion, it was found that there were no initial differences as their average proficiency score was between a B2 and C1 level of English, according to the Common European Framework of Reference for Languages (CEFR), The exception were 3 participants whose proficiency level was in the upper limit of a B1 level, hence been practically equivalent to B2 level. The OPT (Allen, 1992) consists of one-hundred grammar questions, including multiple choice and fill-in-the-gap exercises and assesses the proficiency level of students according to the Common European Framework of Reference for languages (CEFR), which means that test results could range from an A1 level to a C2 level. Given that all participants were in their last year of their undergraduate studies, they had all completed modules on subjects including English language, literature, and Applied Linguistics. Therefore, it can be assumed that, in addition to their intermediate to high proficiency level, as confirmed by the OPT test, they were also knowledgeable in other areas relating to the English language.

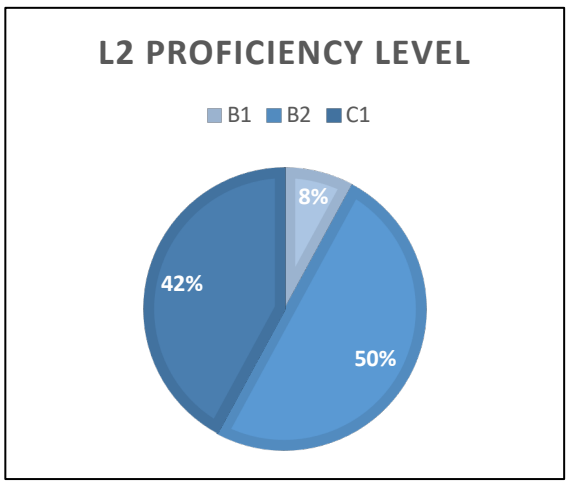


Figure 4. Participants' L2 Proficiency Level According to the CEFR

IV.3. DATA COLLECTION INSTRUMENTS

IV.3.1. The Writing Task

The writing task selected for the study was the complex version of the Fire-chief Task, designed by Gilabert (2005, 2007). This task consists of a problem-solving, picture-based writing activity in which students are presented with an image of a burning building from which a number of people are in need of rescuing (see Figure 5).

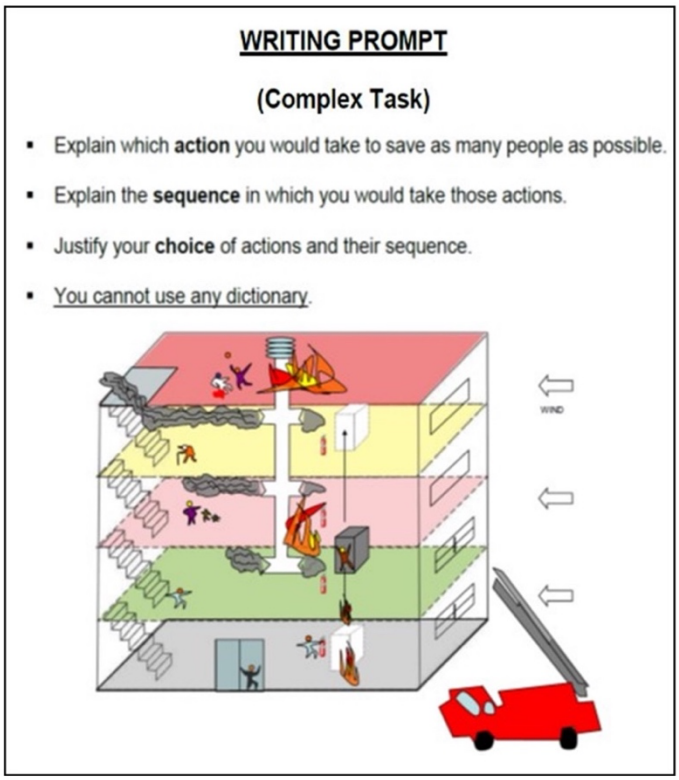


Figure 5. "Fire Chief" Task (adapted from Gilabert, 2005, 2007)

The complex version was created to include people holding specific roles and thus, the decision the students take when completing the task conditions the next decision, and so on until all of the individuals on the graphic have been "saved" (an old man, a pregnant woman, children, etc., are included in the visual prompt, see Appendix 1). Participants were asked to explain (i) the actions they would take in order to save as many people as possible from the burning building, (ii) the order in which they would do so, and (iii) the reasons for which they would take those actions. The fire-chief task was selected as it has been used (and found to be effective) in numerous other studies conducted in the research group to which this thesis is

affiliated. One of the main reasons behind its continued use in the research programme, is the ability the complex version of the task has in eliciting written output (Sánchez et al. 2020). This was particularly important in the thesis due to its´s focus on *writing as a form of language learning*. Providing students with the complex version of the task, as opposed to a simpler one, allowed for an increase in reasoning demands, leading to a higher written output, rendering more opportunities to reflect and process new and old linguistic forms and therefore, leading to a higher chance of consolidation of the target language (Cumming, 1990). As the participants had a high language proficiency, the complex task was selected as a suitable problem-solving task for this level of language competency. The task was measured and validated in a study by Révész, Michel and Gilabert (2016) and the complex version of the fire-chief task was found to elicit more mental effort, with students and teachers confirming that the more complex a task, the greater the cognitive demands. Empirical research on task complexity has also demonstrated that the higher the demands of a task, the more complex the language output (Tavakoli, 2014; Vasylets et al., 2017), particularly concerning accuracy, for students with a higher L2 proficiency (Ishikawa, 2007). As the aim of the thesis was to engage participants in written corrective feedback processing, it was essential to include a task that not only elicited a substantial written output but that also challenged the students in terms of the language needed to complete the task. Therefore, and taking into account the participants higher levels of L2 proficiency, the complex version of the task was selected.

The specific instructions given to participants were that they had a total of 50 minutes to complete the writing task and they were not permitted to use any external help such as dictionaries or online translators (in the case of the digital writing condition).

IV.3.2. WCF Processing Instruments

To answer RQ2 and RQ3, the study explored the methodological affordances of three WCF processing conditions: (i) think-aloud protocols, (ii) written languaging, and (iii) a combination of think-aloud protocols and written languaging in two writing environments (computer-mediated and pen-and-paper writing conditions). Therefore, we used two processing instruments as the third condition entailed a combination of them. The first instrument was a written languaging table designed to guide and encourage students to reflect in writing on the errors made and the feedback provided on them. The second instrument were metacognitive think-aloud protocols, selected to engage students in oral reflection on errors

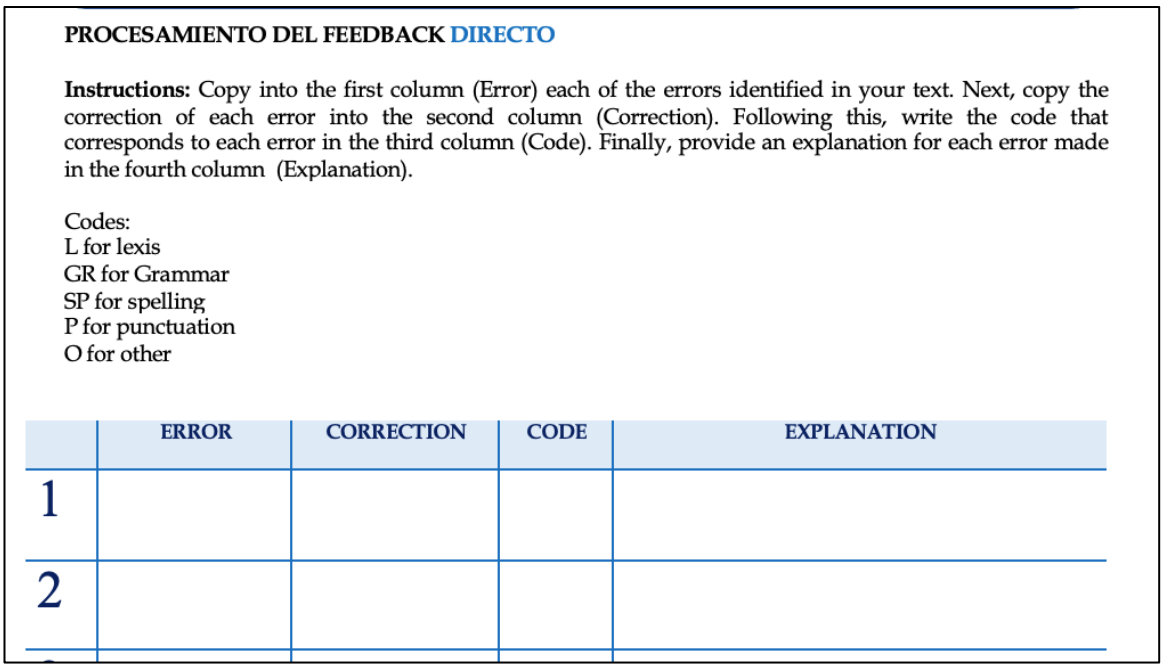
made and feedback received. An in-depth description of both processing instruments is presented below, including relevant information on how the processing instruments were distributed in the two composing environments in focus, i.e., digital, and pen-and-paper writing.

IV.3.2.1. Written Languaging Tables

As discussed in II.3.3.3, the term languaging refers to ¨the process of making meaning and shaping knowledge and experience through language¨ (Swain, 2006, p. 98). Encouraging learners to engage in languaging is purported to provide them with an opportunity to solve language-related issues they may have and, ideally, find solutions to these problems (Suzuki & Storch, 2020). Written languaging consists in providing students with a written outlet in which they can write down their reflections on language-related issues. In feedback processing research specifically, written languaging invites learners to discuss the feedback they have received, including providing explanations for the errors and commenting on the error corrections provided on such errors. Despite receiving less attention than oral languaging in SLA research, written languaging has proven to be a beneficial tool in engaging students in written corrective feedback processing with resulting improvements in L2 written production (e.g., Cerezo et al., 2019; Manchón et al., 2020; Moradian et al., 2020; Simon and Zuniga, 2020). Various tools have been implemented to foster written languaging output, including notetaking (Moradian et al., 2017, 2020), open-ended worksheets (Suzuki, 2012, 2017) and structured written languaging tables (Cerezo et al., 2019 and Manchón et al., 2020). In order to maintain coherence with the research carried out thus far in the research group to which this doctoral dissertation is associated (Cerezo et al., 2019; Manchón et al., 2020), the decision was made to include structured written languaging tables.

The written languaging table used in the study was taken and adapted from Cerezo et al. (2019). The table included four columns corresponding to the four pieces of information participants were asked to provide (see appendices 2 and 3): The first column asked participants to write down the error made in their original text, the second column required them to write down the correction for the error (which they had access to as they received direct WCF), and the third column asked participants to classify the error using the following codes: (i) GR (grammar); (ii) L (lexis); (iii) SP (spelling); (iv) P (punctuation) and (v) O (other). Finally, the last column, invited participants to reflect on the error in terms of a more metalinguistic

analysis through providing an explanation for the error and its correction. Figure 6 shows an extract of the written languaging table used as data collection procedure.



PROCESAMIENTO DEL FEEDBACK DIRECTO

Instructions: Copy into the first column (Error) each of the errors identified in your text. Next, copy the correction of each error into the second column (Correction). Following this, write the code that corresponds to each error in the third column (Code). Finally, provide an explanation for each error made in the fourth column (Explanation).

Codes:
L for lexis
GR for Grammar
SP for spelling
P for punctuation
O for other

	ERROR	CORRECTION	CODE	EXPLANATION
1				
2				

Figure 6. Written Languaging Table

IV.3.2.2. Think-Aloud Protocols

The second data collection instrument used was think-aloud protocols, an online type of verbal report defined by Bowles (2019) as "verbalisations that learners make either while completing a task (think-aloud protocols) or some time thereafter (stimulated recall)" and which "provide information about how learners process the second language" (p. 31). Specifically, think-aloud protocols can be divided according to the type of information they infer, namely metacognitive TAs or nonmetacognitive TAs (as detailed in section II.3.3.3*)*. We opted for metacognitive think-aloud protocols in the study to facilitate comparison with the written languaging tables, which also intended to elicit metacognitive information from participants. In parallel with the instructions provided to participants for the completion of the written languaging tables, the instructions for the think-aloud groups required participants to reflect on their errors and the written corrective feedback received by providing specific reasons and explanations for the errors they had made. The decision behind presenting students with specific instructions for the metacognitive think-aloud protocols was not only to ensure that the information they provided corresponded with the written languaging data, but also due to research finding that providing specific instructions for verbal reports leads to an

improvement in the quality of the data collected. (Cohen, 2013). These instructions were provided in writing (see Figure 7) and orally, as each participant met with the researcher (in person or via *Zoom*) for the think-aloud session. As shown in Figure 7, the written instructions asked participants to speak out loud and say everything that went through their mind whilst processing the error corrections provided on their initial writing. Prior to the individual think-aloud sessions, the researcher spent 5 minutes with the participants, went through the written instructions and complemented them with further indications along the lines of the instructions provided for the written languaging tables. No training was provided in line with the lack of training provided for the written languaging condition.

They were encouraged to think aloud in English but were told that speaking in their L1 was also permitted if they felt more comfortable and, therefore, some of the transcriptions include a mix of both English and Spanish. Giving participants the possibility of using their L1 or their L2 for their think-aloud verbalizations was partly based on the consideration that previous research had found benefits for thinking aloud both in the L1 and the L2 (e.g. Adrada-Rafael & Filgueras-Gómez, 2019).

Following standard procedures, if a student stopped speaking during the think-aloud session, they were reminded that they should continue verbalising their thoughts and they were prompted to keep talking throughout. The researcher was able to do this because, as already mentioned, each participant was recorded one by one in an individual session (either online via zoom or in person, depending on the writing condition), to ensure that there was no background noise that could corrupt the quality of the recordings.

Think Aloud Instructions

In this task I am interested in hearing what you think about when processing your written corrective feedback. So, I am going to ask you to speak aloud and say anything and everything that crosses your mind the entire time you are working. I would like you to talk constantly, without planning what you are going to say. Imagine you are in an empty room, talking to yourself. Remember to keep talking throughout and just say whatever you are thinking. You have a maximum of one hour to complete the task.

Figure 7. Instructions for the Metacognitive Think-aloud Treatment Condition

The participants in the dual written languaging plus think-aloud condition received the same instructions as in those who processed feedback in either the former or the latter condition, as further detailed in IV.4. below.

IV.3.3. Exit Questionnaires

Upon finalizing the data collection, participants were asked to complete an exit questionnaire (see appendix 4). The main aim of the questionnaire was to gain perspective on the participants' perceptions of the direct written corrective feedback they had been provided with and the feedback processing condition and instrument. The questionnaire consisted in three open-ended questions related to how useful the written corrective feedback had been and asked participants whether or not they had been able to incorporate all of the error corrections provided into their re-writing and to provide a reasoning why.

IV.4. DATA COLLECTION PROCEDURES

As shown in Figure 8, the data was collected over four 50-minute sessions, distributed over a total of three days. On day one participants took the proficiency test and on day two they completed the pre-test writing task. On day three the participants were provided with their original texts with direct written corrective feedback. They were invited to reflect on the

feedback in one of the three treatment conditions, namely (i) think-aloud protocols, (ii) written languaging tables and (iii) think-aloud protocols and written languaging tables. In the fourth session (on day three as well), participants were asked to complete the post-test, i.e., to rewrite their original texts on the basis of the feedback received and their processing of it. Upon completing the re-writing, the participants were provided with the exit-questionnaires

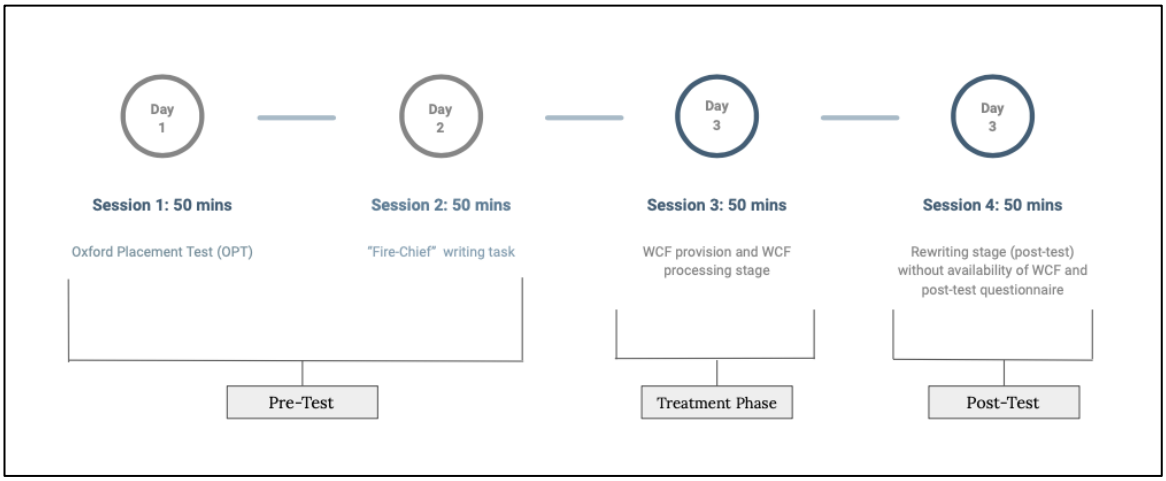


Figure 8. Overview of the Data Collection Procedure

As the participants were divided into two groups according to the writing environment (digital versus pen-and-paper), members of each group completed the data collection in a different manner. In what follows we provide additional details of the data collection procedures in each writing condition, separately.

IV.4.1. Pen-and-Paper Writing Condition

In the first and second sessions, participants in the pen-and-paper writing condition completed the OPT test and the pre-test writing task in a university classroom. The instructions for the writing tasks were those specified in IV.3.1. above (see also Appendix 1). These two actions were completed on pen-and-paper and were submitted by hand to the researcher at the end of each session. The researcher corrected the pre-test writing task by hand, using a marker pen. The third session, which corresponded to the treatment phase of the study, invited students back to the classroom to process the direct written corrective feedback they had been provided with on their written texts. They were divided according to the three processing treatment conditions established, namely, (i) written languaging, (ii) think-aloud protocols and, (iii) simultaneous written languaging and think-aloud protocols.

The participants in the written languaging only condition were grouped together and asked to process their errors via the written languaging table. They were provided with a copy of their written text with direct written corrective feedback (see Appendix 5) and were asked to process their errors by filling out the written languaging table (see Appendix 2). All of this was done using pen-and-paper. The participants in the think-aloud only group were invited to attend an individual session in which they were individually recorded whilst processing their errors following the procedure detailed above in IV.3.2.2.

Finally, the participants who were required to process their errors simultaneously via think-aloud protocols and written languaging, also attended this third session individually, as done by the participants in the think-aloud only condition. As mentioned in IV.3.2.2., in order to ensure a clean recording of the think-aloud protocols, it was essential that each participant attended individually and carried out their feedback processing in a calm, quiet environment. The think-aloud protocols were recorded on two devices (a mobile phone and a laptop) to ensure that the think-aloud verbalizations were successfully recorded and comprehensible. This group of participants completed the written languaging table in the same pen-and-paper conditions as the written languaging only group.

In the fourth and final session, all participants were invited back to a university classroom to re-write their original written texts. They were provided with an original copy of their texts without access to the WCF provided on them (as done in most WCF studies. See for example, Caras, 2019; Cerezo et al., 2019; Manchón et al., 2020) and were required to re-write the text, correcting anything they deemed necessary. By withholding the feedback, we were able to truly compare the effects the WCF processing conditions had had on participants. That is, we were able to observe the incorporations they had made in the rewritings without the aid of the error corrections provided. Upon finalising the re-writing, participants were asked to fill out the exit questionnaires

IV.4.2. Computer-Mediated Writing Condition

The participants in the computer-mediated writing condition followed the exact same data collection procedure as the pen-and-paper writing condition, with the sole difference being that all the data was collected in an online setting. Thus, in the first and second sessions, participants connected to a *Zoom* call. The researcher shared the OPT test via e-mail and the pre-test writing task was shared by giving access to a *Google* document. Data had to be collected online due to the COVID-19 pandemic, given that at the time of lockdown we had

collected only the pen-and-paper data. The participants were fully familiar with the two applications used (*Zoom* and *Google*) as they were the platforms that had been used to the move to online teaching as a result of the pandemic. Therefore, no further training was necessary.

Once the participants had finished their pre-test writing task, a copy was saved on *GoogleDocs* and a second copy was made on which the researcher provided the feedback. In order to ensure the participants did not have access to the documents once they had finished each task, their access was removed.

The third session, corresponding to the WCF processing treatment phase, invited participants to attend a new *Zoom* call. Each participant was provided with access to a *GoogleDocs* which contained their writing plus direct written corrective feedback (see Appendix 6). As in the pen-and-paper condition, the participants in the written languaging only group were grouped together, this time in one *Zoom* session and they filled out the written languaging table online, via *GoogleDocs*. For the two other groups, who were required to think-aloud, each participant was invited to an individual *Zoom* session. This session was recorded using a tool embedded in *Zoom* and the researcher turned off their camera and microphone throughout the process, to ensure that the participant felt at ease whilst thinking aloud. Once the participants had finished, their access to the documents was taken away to ensure that they would not be able to consult them further.

In the fourth and final session, all participants were invited back to attend a *zoom* session and were individually given access to their original texts, without the availability of WCF. They were provided with a new document in which they were asked to re-write their original text, correcting any errors they deemed necessary in order to improve the text. Once participants had finished, their access to both documents was once again taken away and they were given access to the exit questionnaires, which were also completed via *GoogleDocs*.

IV.5. DATA CODING AND ANALYSIS

The data to be analysed were the written texts produced by the participants as pre- and post-tests, and the WCF processing data, namely, the content of the written languaging tables, on the one hand, and think-aloud protocols, on the other. In what follows, the methodological decisions taken in data coding and analysis are described and justified.

IV.5.1. Written Texts

The pre-test and post-test written texts were analysed in terms of CAF (complexity, accuracy, and fluency), as detailed next.

IV.5.1.1. Complexity

Complexity in L2 writing research has been defined as the ability to use advanced language (Ellis, 2008), particularly when related to complex and sophisticated L2 forms and structures (Wolfe-Quintero et al., 1998). The construct of complexity comprises several sub-categories and dimensions that have attracted attention in L2 writing research. Bulté and Housen (2021) proposed a framework which distinguishes between two principal types of complexity: *relative* and *absolute*, in which relative complexity refers to "the difficulty with which linguistic items are learned" (Vasylets, 2017, p. 119) and absolute complexity includes aspects relating to *linguistic*, *discourse-interactional*, and *propositional* complexity (elements which constitute a learner's interlanguage system). Given the focus of this doctoral dissertation on L2 written products, linguistic complexity was chosen as the most relevant dimension of L2 complexity to be analysed and, more specifically, system complexity, which explores complexity "at the level of the language system as a whole" (Vasylets, 2017), as it provides a comprehensive view of a learner's full language production. As the dissertation focuses on L2 written production, two principal subcategories of system complexity were selected for analysis given their relevance for written production and their inclusion as independent variables in many L2 writing empirical studies (Lu, 2010; Ortega, 2003; Wolfe-Quintero et al. 1998): (i) lexical complexity, and (ii) syntactic complexity.

Before calculating the lexical and syntactic complexity of the texts produced by the participants, they had to go through the process of being parsed. Parsing refers to the process of modifying a written text for it to represent syntactic formalism (Pyysalo, 2013). Accordingly, texts were manually transcribed and adapted so that any incorrect lexical items or words and expressions that appeared in the participants L1 were corrected and translated into English (L2). This included any words that were made up or any incorrect spellings. Once the texts had been parsed, they were introduced into Synlex (Ai & Lu, 2010; Lu, 2012), an online tool for analyzing both syntactic and lexical complexity. In response to previous research being criticised for not including a sufficient number of constructs for each type of

complexity (Ellis, 2008; Ortega, 2012), the following sub-constructs were included in the analysis of both lexical and syntactic complexity:

(i) Lexical complexity was analysed by measuring the lexical richness, sophistication, and degree of variety of the L2 participants written texts.
(ii) Syntactic complexity was analysed by looking at five sub-sets, including (i) length of T-Units (MLT); (ii) length of clauses (MLC); (iii) complexity via subordination (DC/C); (iv) complexity via coordination (CP/C): and (v) complex nominals per clause (CN/C)

Lexical complexity was analysed using the free online Lexical Complexity Analyzer (LCA) tool developed by Lu (2010). The following three measures were included in the analysis:

- Lexical richness (LD), which corresponds to the vocabulary size of a language user and, therefore, the higher the number of lexical items found in a text, the higher the lexical richness.
- Lexical sophistication (LS2) refers to the use of vocabulary that is not frequently used in a language, which characterizes language use by more proficient L2 writers (Laufer & Nation, 1995, as cited in Vasylets, 2017). As this measure relates to the use of rare language, the analysis is usually based on frequency counts of words in a language corpus. A high number of infrequent, unusual words in a given text is interpreted as an indication of a high level of lexical sophistication.
- Lexical variety (UBER) concerns the diversity of words used in written (or oral) production. The higher the range of vocabulary used, the higher the lexical diversity of the language user (Vasylets, 2017). This measure therefore takes into account the repetition of words. Therefore, the fewer words that are repeated in a given text, the higher the lexical variety it is said to possess.

Syntactic complexity was analysed taking into consideration general, clausal, and phrasal measures, as proposed by Norris and Ortega (2009). This dimension was analysed using the free, online L2 Syntactic Complexity Analyzer (L2SCA) designed by Lu (2010). The following five categories were included in the analysis:

- Length of production units, analysed via two measures: mean length of clauses (number of words/number of clauses) and mean length of T-Units (number of words/number of T-Units). Clauses are defined by Lu (2010) as any structure which is formed of a subject and a finite verb, "including independent clauses, adjective clauses, adverbial clauses and nominal clauses" (p. 481) whereas T-units are structures that include "one main clause plus any subordinate clause of nonclausal structure that is attached to or embedded in it" (Hunt, 1970, p. 4, as cited in Lu, 2010). These constructs serve to measure overall text complexity.
- Complexity via subordination (DC/C), measured by analysing dependent clause ratio (number of dependent clauses/ total number of clauses)
- Complexity via coordination (CP/C), measured via coordinate phrases per clause (number of coordinate phrases/ number of clauses)
- Focusing on specific structures in the written text, the number of complex nominals per clause were calculated. Complex nominals as defined by Cooper (1976) "comprise (i) nouns plus adjective, possessive, prepositional phrase, relative clause, participle, or appositive, (ii) nominal clauses, and (iii) gerunds and infinitives in subject position" (cited in Lu, 2010, p. 483)

IV.5.1.2. Accuracy

Accuracy refers to the ability to use a language for communication purposes with very minimal (or no) errors (Wolfe-Quintero et al., 1998). Accuracy has been the principal construct used in assessing L2 writing, especially in measuring language gains from pre-test to post-test written output as a function of feedback provision and use. In this study, accuracy was measured as the total number of linguistic errors, divided by total number of words, times 100, a decision taken on the basis of previous L2 writing research (e.g., Chandler, 2003; Manchón et al. 2020; Truscott & Hsu, 2008; Van Beuningen et al., 2012).

In order to calculate the accuracy of the written texts produced by participants, each text was individually corrected by the researcher, all errors were highlighted, and a correction was provided given that unfocused direct WCF was the selected feedback type in the study. Errors were defined as any language use that deviated from the "norm", in comparison to a proficient L2 user.

In terms of reliability measures, 50% of the writing tasks were corrected by another researcher and disagreements were discussed and resolved with an inter-rater reliability of 96% (total number of agreements/total number of errors x 100). The second researcher was an expert Applied Linguistics lecturer and researcher, and member of the research team in charge of the global program of research on L2 writing of which this PhD dissertation is part of. The two researchers met in person and compared the error corrections they had provided on the selected texts. Any disagreements were discussed until a consensus was finally reached. The rest of the analysis was carried out by the PhD candidate.

IV.5.1.3. Fluency

L2 fluency, a multi-dimensional construct that has received ample attention in oral language research, has traditionally been defined as "a person´s general language proficiency" (Housen & Kuiken, 2009). L2 writing fluency in writing has also attracted considerable attention in L2 writing research and has typically been defined as the speed at which a language user can automatically access their L2 knowledge as measured by dimensions including speed fluency, breakdown fluency, and repair fluency (Housen & Kuiken, 2009). Fluency in this PhD thesis has been operationalized according to speed fluency, which refers specifically to the rate at which an L2 learner delivers output.

The participants' writing fluency was therefore measured -following previous research (Wolfe-Quintero et al., 1998)- by calculating the total number of words written per minute (total words/total time spent on task). To measure total time-on-task, the participants were asked to note down the exact times they had commenced and finished writing. These times were corroborated with the ones the researcher has annotated to ensure that there were no discrepancies between the two. All participants completed the task in the 50 minutes stipulated in the task instructions. Given the two writing environments in the study, it was essential to use a measure of fluency valid for paper-based and screen-based writing. Thus, rather than considering time spent on task or number of words written as individual variables, as done in some previous SLA writing research (Vasylets et al., 2017), the decision was made to measure fluency considering the number of words produced in relation to the total time spent on task.

IV.5.2. Feedback Processing Data

The analysis of the feedback processing data was approached from a depth of processing (DoP) perspective. In line with the procedure followed in previous WCF processing studies (Adrada-Rafael & Filgueras-Gómez, 2019; Caras, 2019; Kim & Bowles, 2019), the analysis of the data from the three treatment conditions was based on Leow's (2015) DoP definition, namely:

> The relative amount of cognitive effort, level of analysis, elaboration of intake, together with the usage of prior knowledge, hypothesis testing, and rule formulation employed in decoding and encoding some grammatical or lexical item in the input (p. 204)

In order to guide the data analysis by this operational definition, it was essential to first segment the data into units of analysis that could be categorized according to levels of DoP. However, despite previous research successfully coding the processing data according to levels of DoP, no previous research had explored the three treatment conditions included in this thesis. There was a large variation in the data output provided by each feedback processing condition, as evidenced in the three excerpts below corresponding to the annotations from the written languaging (WL) tables, verbalisations from the think-aloud (TA) protocols and the combination of both in the WL+TA. Therefore, it became extremely difficult to code the data according to the same criteria. This variation in data output led us to approach the data coding individually, analysing the WL and TA data separately, and to subsequently elaborate a common coding scheme that was applicable to the data in the three treatment conditions. These analyses are detailed in what follows.

[1] Think-aloud data output

TA data: Original	TA data: Translation
[1] [WHISPERS: would open the main door so that the people that are inside the building] … el de could…hmm…yo (2) bueno… I always (2) make this mistake…and (2) I don't know…I just don't…es como que no lo interiorizo ese…ese tiempo allí pero bueno (3) are trapped inside the building (2) can exit (2) vale…si…si porque como el are está en presente (2) pues el (2) can…o sea el próximo verbo también tiene que ir en el mismo tiempo verbal (2)	**[1]** WHISPERS: would open the main door so that the people that are inside the building]… el de could…hmm…I (2) well…I always (2) make this mistake…and (2) I don't know…I just don't know…it is like I don't interiorize this…this tense there but yeah (3) are trapped inside the building (2) can exit (2) ok…yes…yes because the are is in present (2) so the (2) can…I mean the next verb also has to be in the same verb tense (2)

[2] Written languaging data output

WL data:

[2]

Error	Correction	Code	Explanation
Fog	Smoke	V	"Fog" is a meteorological phenomenon, "smoke" is the correct word
Estinguished	Extinguished	SP	A spelling mistake, perhaps triggered by the influence of the Spanish "extinguir" which is usually pronounced as "estinguir"

[3] Written languaging and think-aloud data output

WL & TA data:

[3] WL

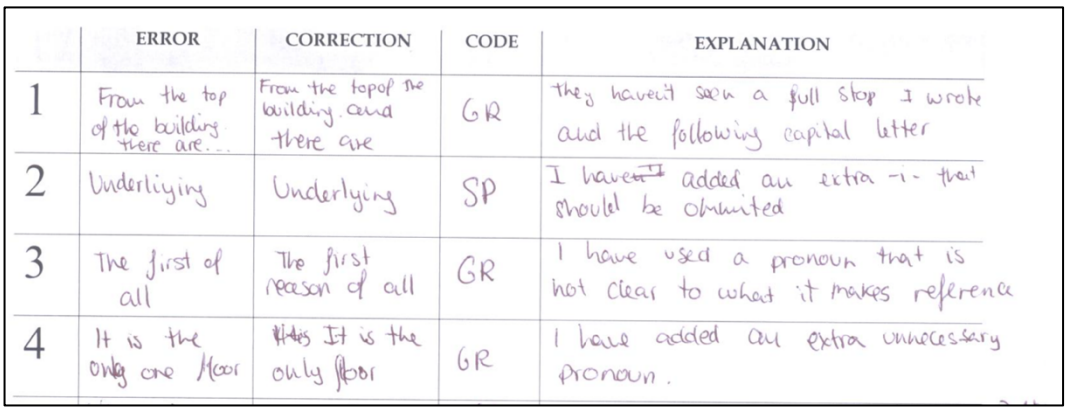
Error	Correction	Code	Explanation
Origen	Source	V	The second term is more accurate

TA data: Original	**TA data: Translation**
[4] Vale…as they have the origin of the fire…the source (2) porque origen sería mas…vale…creo que origen (3) y source (2) al pensarlo en español creo que aqui en inglés hay una division que en español no lo tenemos [filling in the WL table] origin (2) source (6) porque no son…en español distinguimos (2) no…solo está origen y aquí está origen y source (2) esto también es vocabulario (3) [humming] (9) hmm… el segundo término…the second term [filling in the WL table] is… more…accurate (2)	**[4]** Ok…as they have the origin of the fire…the source (2) because origin would be…ok…I think that origin (3) and source (2) thinking about it in Spanish I think that here in English there is a distinction that in Spanish we do not have [filling in the WL table] origin (2) source (6) because they are not…in Spanish we disringuish (2) no…there is only origin and here we have origin and source (2) this is also vocabulary (3) [humming] (9) hmm…the second term…the second term [filling in the WL table] is…more…accurate (2)

IV.5.2.1. Written Languaging Tables

The coding scheme in Cerezo et al.'s (2019) study -based on Leow's (2015) operationalization of DoP- was adopted for the analysis of the WL data. This coding scheme had been implemented for WL data in the studies carried out within the global program of research of which this dissertation is a part of. The coding scheme distinguishes between three main levels of awareness, namely, noticing, reporting, and understanding.

Upon application and reiterating the findings in Manchón et al. (2020), it quickly became evident that the data output provided by the WL tables did not relate to the processing stage itself but, rather, it was a product of the outcome of the feedback processing stage. Thus, the WL tables provided evidence on metalinguistic information, rather than on levels of processing (an example can be seen in Figure 9).



	ERROR	CORRECTION	CODE	EXPLANATION
1	From the top of the building there are...	From the topof the building and there are	GR	they haven't seen a full stop I wrote and the following capital letter
2	Underliying	Underlying	SP	I have~~n~~ added an extra -i- that should be ommited
3	The first of all	The first reason of all	GR	I have used a pronoun that is not clear to what it makes reference
4	It is the ~~Only~~ one floor	~~It~~ It is the only floor	GR	I have added an extra unnecessary pronoun.

Figure 9. Extract of a Written Languaging Table.

Taking into consideration these two factors, the following steps were taken in order to analyse the WL data:

Step 1:

The analysis of the WL tables consisted in first approaching the data from a noticing perspective, that is, categorising whether the errors had been noticed or unnoticed, as evidenced in whether or not the error has been included in the first column of the WL tables.

Step 2:

In cases where errors were noted down, these were further categorised according to the amount of information participants had included in the columns of the WL tables corresponding to the correction, code, and explanation. Therefore, as seen in Table 6, and following Cerezo et al's (2019) coding scheme, verbalizations on noticed errors were divided into five sub-categories, which corresponded to the amount of information included in the table:

1. Level 1 (awareness at the level of noticing): Participants only providing the error transcription.
2. Levels 2 and 3 (awareness at the level of reporting): Participants provided not only the error transcription but also either the error correction, the error category (Level 2) or both (Level 3).
3. Levels 4 and 5 (awareness at the level of understanding): Participants provided (i) the error transcription and a metalinguistic explanation, (ii) the error transcription, error correction and metalinguistic explanation, (iii) the error transcription, error category

and metalinguistic explanation or, all four elements, (iv) error transcription, error correction, error category, and metalinguistic explanation (level 5).

Table 6. Coding Scheme Adopted for the Analysis of the Written Languaging Tables (Source: Cerezo et al., 2019)

Awareness at the level of noticing	Level 1:	Error transcription alone
Awareness at the level of reporting	Level 2:	Error transcription plus either error correction or error category
	Level 3:	Error transcription, error correction, and error category
Awareness at the level of understanding	Level 4:	Error transcription and metalinguistic explanation OR Error transcription, error correction, and metalinguistic explanation OR Error transcription, error category, and metalinguistic explanation
	Level 5:	Error transcription, error correction, error category, and metalinguistic explanation

All the WL data was coded by the main researcher and author of this doctoral thesis. An intra-rater reliability score of 100% was calculated when the researcher re-coded a sample of 20% of the WL data, proving no discrepancies between the first and second coding (conducted two weeks apart) hence providing a more than satisfactory intra-rater reliability score.

IV.5.2.2. Think Aloud Protocols

The metacognitive think-aloud protocols (recorded, as mentioned in an earlier section using both the Voice Memo application on the researcher's computer as well as on a mobile phone device) data was transcribed by the researcher and a *Word* document for each participant was created, in preparation for coding. As done with the written languaging tables, the first

step was to analyse the data according to depth of processing. In this case, we again wanted to follow Leow's (2015) DoP operationalization. To this end, the TA data was segmented in terms of language-related episodes (LREs) as done in previous oral verbalisation research (e.g., Swain & Lapkin, 2002; Wigglesworth & Storch, 2012). An LRE is a unit of analysis for TAs that Swain & Lapkin (1995) referred to as "any segment of the protocol in which a learner either spoke about a language problem he/she encountered while writing and solved it either correctly or incorrectly or, simply solved it without having explicitly identified it as a problem" (p. 378). Given the metacognitive nature of the TAs (given that the participants were explicitly asked to comment on the errors made and corrections provided on their written texts), the identification of the LREs was relatively straightforward. Therefore, every time a participant made a comment on an error and its correction, the LRE was identified and coded. For the coding of the LREs, a data-driven process was followed in which each episode was coded according to the actions implemented by the students upon processing the error corrections. As a result (see Figure 10), a list of broad categories was established, which simply tried to capture all the actions evidenced in the participants' verbalizations.

- Read/repeat target structure
- Disagrees with error correction (no explanation given)
- Disagrees with error correction (explanation given)
- Fully understands the error correction and provides an accurate explanation
    - Reason is L1 translation
    - Rule formulation
- Attempts but incomplete/incorrect explanation
- Accepts the correction as they didn't know how to say it in their L2
- Cannot explain the error, isn't sure of the mistake
- Ignores the error
- Accepts the error as it was made due to rushing
- Understands, "always makes this error", provides an (in)complete explanation
- Translates the error to the L1 and accepts with little/no explanation

Figure 10. Broad Categories for the LRE's (Source: McBride & Manchón, 2023)

Informed by Leow's (2015) DoP definition, the data-driven coding led the researcher to first establish levels of engagement with the WCF in as similar way as possible to the WL data. As a result, two macro-groups were established corresponding to whether or not the students engaged with the error corrections or not, as seen in Figure 11.

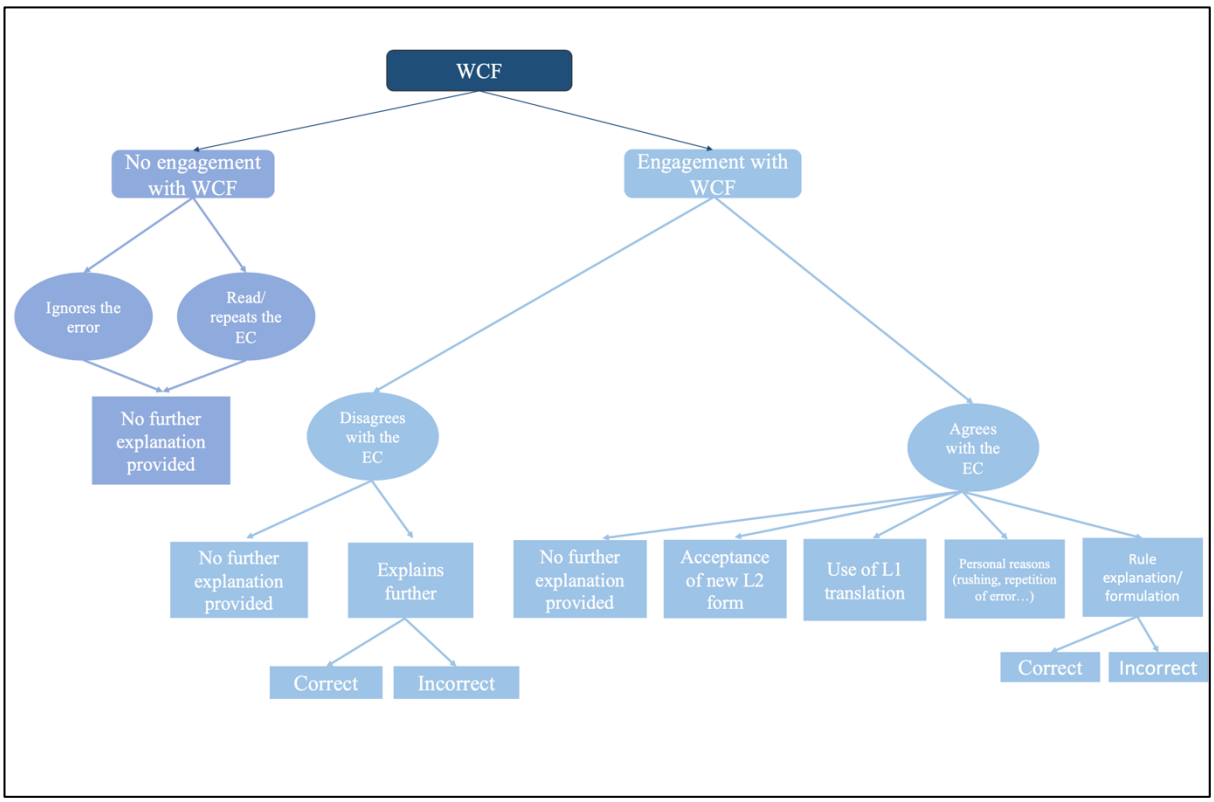


Figure 11. Coding of TA Data According to Engagement with WCF (Source: McBride & Manchón, 2023)

Once the metacognitive TAs had been coded according to the participant's engagement (or lack thereof) with the WCF, the following step was to establish levels of depth of processing. As mentioned in the literature review (see especially II.3.3.3), the coding of DoP has been approached in a variety of ways in previous studies. For some authors, LREs were coded according to two levels, namely, high and low (Kim & Bowles, 2019) whereas others distinguished three levels, namely, deep, intermediate, and low (Adrada-Rafael & Filgueras-Gómez, 2019), or high, medium, and low (Caras, 2019). Therefore, taking these classifications into account, further steps were taken in order to identify the levels of DoP of the TA data.

Step 1:

The initial broad categories, corresponding to whether or not participants had engaged with the WCF, were refined according to high and low levels of DoP (following Kim & Bowles, 2019). In their study, high DoP was characterised by participants spending a long-time processing WCF, showing high cognitive effort, and formulating target rules. In contrast, low DoP corresponded to participants simply recognising a form by reading/repeating it, minimal cognitive effort, and a general lack of understanding of the correction provided.

Upon coding the data for the study, the data analysis was further refined by distinguishing three levels of DoP: high, medium, and low. The decision to include three levels was a result of some TA verbalisations not fitting the criteria for high and low (as in Kim & Bowles, 2019) and thus, a third level was included corresponding to medium DoP, as done in Caras (2019).

Step 2:

The following step entailed categorising the LREs according to the three DoP levels established and taking into consideration the actions observed (as described in Figure 10). However, when coding the data, we observed that other aspects had to be taken into consideration, following Leow (2015). We are referring to criteria such as the cognitive effort expended (measured in terms of time spent on engaging with the corrections and overall effort when successfully or unsuccessfully explaining the errors), and rule formulation attempts and reference to prior knowledge. As a result, the following criteria were established for each DoP level:

1. **High DoP** corresponded to participants:
    i. successfully formulating rules;
    ii. formulating hypotheses about the error corrections;
    iii. successfully translating the error corrections and providing a correct metalinguistic explanation;
    iv. demonstrating high levels of cognitive effort, as manifested in the time (measured in seconds) spent engaging with the error corrections and providing correct metalinguistic explanations;
    v. disagreeing with the correction given but providing a correct metalinguistic explanation to justify their disagreement.

Example 4 of high DoP below shows the participant's hypotheses about the different usages of the word they had initially used (origin) and the correct form (source). By engaging with the target form, they are demonstrating high levels of cognitive effort which results in the student comparing the error correction to their L1 (Spanish) and concluding that the term provided in the WCF is more adequate.

[4]

Metacognitive TA LRE: Original	**Metacognitive TA LRE: Translation**
Vale…as they have the origin of the fire… the source (2) porque origen sería más…vale…creo que origen (3) y source (2) al pensarlo en español creo que aquí en ingles hay una división que en español no lo tenemos [Filling in the table] origin (2) source (6) porque no son…en español distinguimos (2) no…solo está origen y aquí está origen y source (2) esto también es vocabulario (3) [humming] (9) Hmm…el segundo término… the second term [filling in the table] is…more…accurate (2)	Ok…as they have the origin of the fire…the source (2) because origin would be…ok…I think that origin (3) and source (2) thinking about it in Spanish I think that here in English there is a distinction that in Spanish we do not have [filling in the WL table] origin (2) source (6) because they are not…in Spanish we distinguish (2) no…there is only origin and here we have origin and source (2) this is also vocabulary (3) [humming] (9) hmm…the second term…the second term [filling in the WL table] is…more…accurate (2)

2. **Medium DoP** corresponded to participants:

i. translating the error correction (successfully or unsuccessfully) but not providing a metalinguistic explanation or any further information.
ii. discussing having made the same mistake previously in the text (repeated error) and providing a correct explanation.
iii. attempting to explain the error correction metalinguistically but in terms of cognitive effort, very little time is spent engaging with the target item and/or the attempt to explain was abandoned.

In example 5 below, the participant recognises the error correction and successfully states that the incorrect tense has been used. However, no further explanation is provided as to why one tense is more appropriate than the other and the participant simply notes down the error in the table and moves on to the next error correction.

[5]

Metacognitive TA LRE: Original	Metacognitive TA LRE: Translation
A ver, (2) after that, I will save (2) aquí es que de repente me he cambiado de hmm temporal (2) hmm (1) will sería would y esto es fallo de gramática (2) [filling in the table] hmm (2) conditional (1) is needed [writing it down] hmm a ver (10)	Let's see (2) after that, I will sabe (2) here all of a sudden I changed hmm tense (2) hmm (1) will would be would and this is a grammatical mistake (2) [fills in the WL table] hmm (2) conditional is needed hmm let's see (10)

3. **Low DoP** corresponded to participants:

i. reading or repeating the error correction;
ii. simply agreeing (or disagreeing) with the correction provided and not providing any further information;
iii. not understanding the correction provided (as manifested by signs or verbalisations).

In the following example of an LRE categorised as low DoP (example 6), the participant simply reads and repeats the error correction provided, agrees with the correction ("it's more appropriate") and moves on to the next error.

[6]

Metacognitive TA LRE: Original
The next one is about another modal verb (3) I wrote could instead of can (2) and I think it's…it's more appropriate [writes down the error, correction, code and explanation (24)]

When applying this coding scheme to the metacognitive TA verbalisations, some LREs did not fit into the three levels established. These verbalisations (or, in this case, lack thereof) corresponded to instances in which a participant ignored the correction provided, as manifested in leaving all columns blank in the WL table and not mentioning the error at all in the TA protocol. Therefore, in the absence of any recognition of the error correction, these instances were coded as null. As shown in Table 7, the final coding scheme for the conceptualization of DoP included four levels: high, medium, low, and null.

Table 7. Final Coding Scheme for the Levels of DoP of Metacognitive TAs

	Null DoP	Low DoP	Medium DoP	High DoP
Criteria	No cognitive effort and no time spent on processing target form	Low cognitive effort and minimal time spent on processing target form	Target form is commented on but with little discussion/attempt to provide an explanation. Brief engagement with target form.	High cognitive effort, accurate rule formulation, high engagement/ time spent on each target item
Examples	Avoids the error correction or ignores it	Reads the feedback. Repeats the target form. Spends very little time on the correction. Uses words such as yes, ok, hmm before moving on quickly.	Translates into L1 (correctly/incorrectly) but quickly moves on. Recognises making the same error previously. Attempt at providing an explanation/ discussion of error but very brief (correct/incorrect).	Hypothesising about the target form. Translating the target from to L1 correctly and providing an explanation for the error. High cognitive effort as seen in time spent on processing. Correct rule-formation. Disagrees with the correction but provides a correct metalinguistic explanation.

All of the metacognitive data was transcribed and coded by the main researcher and author of this doctoral thesis. A strong intra-rater reliability score of 86% was calculated when the researcher re-coded a sample of 20% of the total data.

IV.5.2.3.WCF Processing Data: Final Coding Scheme

After analysing each introspective measure separately, a coding scheme that could serve for both instruments (i.e., written languaging tables and TA data) was essential to compare the data output each instrument provided and hence validly answer our second and third research questions. One of the main challenges regarding the creation of a comparable coding scheme was the consequence of the different nature of the data output obtained by each instrument. As seen in the previous sections, the data from the written languaging tables was coded according to degrees of awareness, whereas the data from the metacognitive TAs was coded according to levels of depth of processing. This was done in this manner because, as discussed in the literature review, and reiterating findings in Manchón et al. (2020), rather than the WL providing information on the levels of depth of processing, the nature of the data elicited from the WL tables was more related to the outcome of the processing stage, representing the amount of metalinguistic information each participant had provided. Thus, in order to compare the coding of the TA verbalisations and the WL tables, the following steps were taken:

Step 1:

We first established a correspondence between the participants engagement with the WCF provided in the WL tables and TA data. By doing so, we were able to create a table to guide subsequent comparative analyses in which the data output from both the WL tables and the metacognitive TA protocols could be equated as seen in Table 8.

Step 2:

The following step was to reflect the way in which the participants engaged or not with the WCF, as manifested in their TA protocols or the WL table data. Therefore, for the TA protocols, no engagement with the feedback corresponded to a participant ignoring the error or simply reading/repeating the error with no further comment, whereas in the WL data, no engagement with the WCF corresponded to a blank space in the first column of the table (error correction column) or an annotation of the error and correction but no further information provided in the remaining columns.

Table 8. Coding Scheme Comparing WL and TA WCF Processing

	TA	WL
(i) Participant does not engage with the EC	Ignores the error	1. Leaves a blank space
	Reads/Repeats the error	2. Annotates the error and correction with no further analysis
(ii) Participant engages with the EC		
a. Disagrees with the EC	States they disagree but no further verbal explanation	1. The participant states "I disagree" in the explanation section, but no further information provided.
	Explains further by providing a justification/explanation for their use of the term that has been corrected	2. Provides a written explanation for their disagreement
b. Agrees with the error correction	No further explanation/cannot provide an explanation	1. Explanation section left blank or evidence of not being able to provide an explanation
	Acceptance of new L2 form	2. Evidence in the explanation section that a new L2 form has been accepted. "I didn't know that word/form"/ "I didn't know how to say it"
	Use of L1 translation	3. Translation included in the explanation section
	Personal reasons (i.e., rushing, always make this mistake.)	4. Personal reasons (i.e., rushing, always make this mistake.)
	Rule explanation/Formulation	5. Rule explanation/Formulation included in the explanation section

As seen in Table 8, instances of the participants' engagement with the error corrections were divided into two categories (i) instances when participants disagreed with the correction, and (ii) instances when the participants agreed.

When participants disagreed with the correction provided, they did so by stating that they disagreed and, in some cases, by providing an explanation of why. Examples of these cases will be provided in the Results section. On the other hand, when a participant agreed with the correction, there were five main actions identified, including (i) not providing any further explanations or not being able to; (ii) accepting a new L2 form not previously known; (iii) using their L1 to translate; (iv) providing personal reasons for the error and (v) explaining or formulating a rule. Table 8 shows the correspondence of these five actions as well as how they were manifested in both the TA and WL data.

All of the data was coded by the main researcher and author of this PhD thesis. An intra-rater reliability score of 74% was obtained when the researcher re-coded a sample of 20% of the total data. Discrepancies were identified and decisions were taken to continue with reliable data analysis. The resulting coding scheme was applied to the data analysis, which proved to proceed without difficulty.

IV.5.3. Exit Questionnaire

The exit questionnaire (see Appendix 4) was analysed primarily from a qualitative perspective in order to gain insights into the participants' views of the feedback processing stage. Thus, the actions taken by the participants in the processing stage were compared to their responses in the three open-ended questions included in the questionnaires to gain a deeper understanding of how participants approached the processing task. Questions included whether or not they found the feedback provided useful, whether they had been able to incorporate the error corrections in the rewriting, and if they had not revised the text according to the feedback provided, they were asked to provide reasons why.

IV.5.4. Quantitative and Qualitative Analyses

Within-groups (i.e., within written corrective feedback conditions) and between-groups (i.e., digital versus pen-and-paper writing environments) analyses were performed on the quantitative data, which corresponded to the written products (CAF measures), the data on the levels of engagement (WL tables) and the data on levels of DoP (TA protocols). Descriptive measures were calculated using JASP 0.14.1. The decision to include a descriptive analysis rather than a more sophisticated one was based on the consideration that the study included a

wide range of variables (corresponding to the processing instruments and writing conditions) yet a rather small sample of participants.

For the analyses related to RQ1 and RQ2, the means and standard deviations for each CAF measure are obtained. In addition, and in order to establish the magnitude of the differences found among the variables analysed, effect sizes were calculated, following the benchmarks established by Plonsky et al. (2021) for L2 research. Therefore, effect sizes were interpreted according to the following parameters:

Table 9. SLA Research Effect Size (d) Benchmarks (Plonsky et al., 2021)

Effect Size	Small	Medium	Large
Between-groups	0.40	0.70	1.00
Within-groups	0.60	1.00	1.40

The decision to include effect sizes was due to the relevance this parameter has gained in SLA research over the past few years. In addition, due to the lower number of participants involved in the study and the large number of variables included in the research design, effect sizes were able to provide a well-constructed and reliable interpretation.

In terms of RQ3, descriptive results were calculated based on the coded WL tables and metacognitive TAs. The levels of awareness and engagement coded in the WL tables were analysed via means and standard deviations. Qualitative analyses were also undertaken to explore the specific ways in which the participants approached the processing task. For the metacognitive TAs, the levels of DoP were analysed, again via mean and standard deviations, and a more qualitative analysis was conducted in order to explore the participants' cognitive processing. Finally, a qualitative analysis of the exit questionnaire was conducted in order to compare the participants' perceptions of their experience in the treatment conditions and to shed light on whether or not being provided with a WL table whilst processing had any influence on the way in which participants processed their errors while thinking aloud.

CHAPTER V. RESULTS

This chapter will report the data in terms of the answer provided to the 3 research questions guiding the study, which are reproduced next to facilitate the reader's task:

RQ.1 How does writing in a traditional pen-and-paper environment versus writing in a computer-mediated environment affect L2 written production in terms of CAF measures?

RQ. 2 How does the experimental manipulation during WCF processing affect L2 written production (in terms of CAF measures) in pen-and-paper versus computer-mediated writing environments?

RQ.3. How does the experimental manipulation during WCF processing affect L2 learners' levels of depth of processing of the feedback received in pen-and-paper versus computer-mediated writing environments?

IV.1. RESEARCH QUESTION ONE (RQ1). EFFECT OF WRITING ENVIRONMENTS ON L2 TEXTS CHARACTERISTICS

The first research question addressed how writing in a traditional pen-and-paper (P&P) environment versus writing in a digital environment (DW) may affect L2 written production. Given that the participants' L2 written texts were analysed in terms of CAF measures (see Chapter IV), we next report the impact the writing environment had on each CAF component of written production in the pre-test writing task. To validly inspect the sole effect of the variable of writing environment on text characteristics (not in combination with the availability of WCF), only the initial texts provide the valid data for the first research question. The texts written as post-test were written after receiving and processing feedback. Table 10 shows the descriptive data for the effects of writing environment in written products.

Table 10. CAF Measures for Writing 1 in P&P and DW Environments

		Pen-and-Paper Writing		Computer-Mediated Writing		
		Pre-Test		Pre-Test		
		$\bar{X}$	SD	$\bar{X}$	SD	Cohen's d Effect Size
Accuracy	**A**	10.58	3.77	5.88	3.03	**-1.374**
Fluency	**F**	10.30	2.96	12.97	5.97	**0.567**
Lexical Complexity	**LD**	0.45	0.04	0.44	0.02	**-0.548**
	LS2	0.17	0.04	0.17	0.04	0.142
	UBER	15.95	1.09	15.93	1.99	-0.002
Syntactic Complexity	**MLT**	0.04	0.01	0.04	0.01	**-0.558**
	MLC	0.09	0.01	0.08	0.02	**-0.530**
	DC/C	0.44	0.10	0.50	0.06	**0.794**
	CP/C	0.25	0.10	0.25	0.08	0.106
	CN/C	1.16	0.33	1.04	0.17	**-0.449**

In terms of **accuracy** (operationalised as total number of errors/total number of words *100), the texts written in the DW condition were more accurate. We can observe that the participants in the DW group made considerably less errors when compared to the P&P writing condition (5.88 and 10.58, respectively), with a large effect size (d = -1.374). As will be more fully elaborated in the Discussion, participants in the digital writing group were writing in *GoogleDocs*, therefore their initial quantity of spelling and punctuation errors was minimal most likely due to the availability of spellcheck for this writing condition.

In relation to **fluency** (as measured by the number of words produced per minute), the digital writing condition, as expected due to keyboarding, resulted in a slightly higher average

number of words per minute (12.97) when compared to the paper-based writing condition (10.30), with a small effect size (d = 0.567).

Complexity, as detailed in Chapter IV, included measures of both lexical and syntactic complexity. As shown in Table 10, minimal differences were found between the two writing environment conditions, with participants in both writing conditions performing almost identically. Thus, for lexical measures, including lexical sophistication (LS2) and lexical variety (UBER), no differences were found. Similarly, for syntactic measures, including mean length of T-units (MLT) and complexity via coordination (CP/C), results were the same for both writing conditions.

For the sake of comprehensiveness, some minor differences might be mentioned. In terms of lexical complexity, we can observe that the participants in the P&P writing condition performed slightly better than the DW group in terms of lexical density (0.45 and 0.44, respectively), with a small effect size (d = -0.548). Regarding syntactic complexity, the P&P group again performed slightly better than the DW group for mean length of clauses, (0.09 and 0.08, respectively), with a small effect size (d = -0.530), as well as for complex nominals per clause (1.16 and 1.04, respectively) with a small effect size (d = -0.449) On the other hand, it was the DW group who outperformed the P&P condition with regards to the number of dependent clauses (0.50 and 0.44, respectively), with a medium effect size (d = 0.794).

In short, the answer to our first research question can be synthesised as follows:

- The DW condition performed better in terms of accuracy, fluency, and, to a lesser extent, one dimension of syntactic complexity (number of dependent clauses).
- Although the differences were minimal, the P&P writing condition performed slightly better in terms of lexical density, and 2 dimensions of syntactic complexity (mean length of clauses and complex nominals per clause).

V.2. RESEARCH QUESTION TWO (RQ2). EFFECT OF PROCESSING CONDITIONS ON WRITTEN TEXTS ACROSS WRITING ENVIRONMENTS

The second research question asked whether the manipulation of WCF processing conditions had an effect on the participants´ L2 written production, and whether such effects varied as a function of the writing environment. Accordingly, results will be discussed by looking at the effects of the three treatment conditions, namely, (i) think-aloud protocols, (ii) written languaging, and (iii) the simultaneous action of thinking aloud whilst languaging in written form, on the characteristics of the texts written as pre- and post-tests (as measured by CAF indices). Additionally, we will explore whether the variable of writing environment (P&P and DW) mediated the results observed. In order to facilitate the discussion of the results for RQ2, each measure of L2 written production (CAF) will be discussed separately.

V.2.1 Accuracy

In terms of changes in accuracy (operationalised as total number of errors/total number of words*100), from Time 1 to Time 2, Table 11 shows that the texts written as post-test were more accurate for all feedback processing conditions, in the two writing environments.

Table 11. Accuracy Measures at Time 1 to Time 2 across Writing Environments and for All WCF Conditions

		Accuracy Time 1 $\bar{X}$ (SD)	Accuracy Time 2 $\bar{X}$ (SD)	Cohen's d ES
Pen-and-Paper Writing	TA	9.13 (3.30)	2.94 (1.88)	**2.349**
	WL	11.64 (5.34)	3.07 (2.79)	**3.013**
	WL+TA	10.98 (2.15)	1.86 (0.68)	**4.916**
Computer-Mediated Writing	TA	5.41 (3.71)	1.78 (1.37)	**1.502**
	WL	5.70 (3.74)	1.45 (1.12)	**1.573**
	WL+TA	6.53 (1.60)	2.05 (0.82)	**3.901**

As can be observed, the think-aloud plus written languaging condition was the group who improved the most in terms of error reduction from Time 1 to Time 2. In the P&P condition, the WL+TA group experimented a decrease in global errors from Time 1 (10.98) to Time 2 (1.86), with a reduction of 9.12 and a large effect size (*d* =4.916). In the digital writing condition, the WL+TA group reduced their global error percentage from 6.53 at Time 1 to 2.05 at Time 2 (a reduction of 4.48, again with a large effect size (*d* =3.901).

The texts written in the written languaging only condition also showed reductions in overall errors. In the P&P condition, the global percentage of errors decreased from 11.64 at Time 1 to 3.07 at Time 2 (a reduction of 8.57), with a large effect size (*d* = 3.013), whereas in the DW condition, errors were reduced from 5.70 at Time 1 to 1.45 at Time 2 (a reduction of 4.25), again with a large effect size (*d* = 1.573).

Finally, the think-aloud only condition showed the lowest amount of global error reduction, albeit still notable with a decrease in errors from 9.13 at Time 1 to 2.94 at Time 2 for the P&P condition (a reduction of 6.19), with a large effect size (*d* = 2.349). On the other hand, the TA condition in the DW group reduced their global percentage of errors from 5.41 at Time 1 to 1.78 at Time 2 (a reduction of 3.63), once again with a large effect size (*d* = 1.502).

In short, regardless of the WCF processing conditions, all treatment groups showed a reduction in global errors, with large effect sizes. Additionally, the TA+WL group showed the largest reduction in errors in both P&P and DW writing conditions, followed by the WL only group and, lastly, by the TA only group.

V.2.2. Fluency

The second CAF dimension analysed was fluency, measured as number of words produced per minute. Table 12 shows the within-group descriptive data for each WCF treatment condition. As with accuracy, all three treatment conditions resulted in fluency improvements as participants were found to produce more words per minute in the re-writing of their original texts.

Table 12. Fluency Measures from Time 1 to Time 2 for all Processing Conditions

		Fluency Time 1 $\bar{X}$ (SD)	Fluency Time 2 $\bar{X}$ (SD)	Cohen's d ES
	TA	9.76 (3.3)	16.68 (2.99)	**-2.349**

Pen-and Paper Writing	WL	9.53 (2.23)	18.97 (3.40)	**-2.515**
	WL+TA	11.62 (3.22)	19.62 (3.40)	**-2.027**
Computer-Mediated Writing	TA	13.13 (6.73)	21.24 (12.53)	**-1.153**
	WL	13.33 (6.23)	29.30 (11.59)	**-1.884**
	WL+TA	12.43 (6.03)	22.58 (12.22)	**-1.18**

Focusing specifically on each WCF processing condition, the texts written in the think-aloud only group were the ones showing the lowest increase in overall fluency. In the P&P condition, 9.76 words per minute at Time 1 increased by 6.92 to a total of 16.68 at Time 2 (with a large effect size of 2.349) and in the DW condition, 13.13 words per minute at Time 1 increased by 8.11 to 21.24 words per minute at Time 2 (with a large effect size of -1.153). The texts written in the simultaneous WL+TA, P&P condition showed an increase of 7.99 words per minute (from 11.62 at Time 1 to 19.62 at Time 2) with a large effect size (d = -2.027). In the DW condition, the texts written by the WL+TA group increased their fluency from 12.43 words per minute at Time 1 to 22.58 at Time 2, experiencing an increase of 10.33 words per minute, with a large effect size (d = -1.18).

The WCF processing group who increased their fluency the most was the WL only group. In the P&P condition, this group experienced an increase of 9.44 words per minute (from 9.53 at Time 1 to 18.97 at Time 2) with a large effect size (d = -2.515). In the DW condition, the WL group increased their fluency from 13.33 words per minute at Time 2 to 29.30 words per minute at Time 2 (an increase of 15.98 words per minute) again with a large effect size (d = -1.884).

In short, regardless of WCF processing treatment conditions the texts written at Time 2 showed an increase in writing fluency, with large effect sizes. Additionally, the WL only condition showed the highest increase, followed by the TA+WL group and, finally, the TA only condition.

V.2.3. Complexity

The final dimension of L2 written production measured was complexity. As shown in Table 13, L2 writing was rated in terms of its lexical sophistication (LS2), lexical density (LD) and lexical diversity (UBER), as well as its syntactic complexity, including, mean length of t-

units (MLT), mean length of clauses (MLC), coordinate phrases per clause (CP/C), dependent clauses per clause (DC/C) and complex nominals per clause (CN/C). With the objective of facilitating the legibility of the analysis, the results will be divided into two sections, first according to lexical complexity and, second, according to syntactic complexity.

Table 13. Complexity Measures Included in Analysis

Complexity Measures	
Lexical Complexity	
Lexical Sophistication	LS2
Lexical Density	LD
Lexical Diversity	UBER
Syntactic Complexity	
Mean length of t-units	MLT
Mean length of clauses	MLC
Coordinate phrases per clause	CP/C
Dependent clauses per clause	DC/C
Complex nominals per clause	CN/C

- *Lexical Complexity*

The lexical complexity dimension of L2 writing production was analysed in terms of three indices, namely, lexical density (LD), lexical sophistication (LS2), and lexical diversity (UBER).

Regarding lexical density, that is, the lexical richness of the written texts, minimal differences were observed from Time 1 to Time 2 regardless of the WCF processing treatment condition although noteworthy effect sizes can be observed. The think-aloud only group evidenced a minimal decrease from Time 1 (0.47) to Time 2 (0.46) with a medium effect size (d = 0.913) in the P&P writing condition. However, in the DW condition, there was an increase from Time 1 (0.43) to Time 2 (0.44) with a large effect size (d = -1.291) for the TA group. This increase can also be seen for the WL group in the DW condition, who also improved their score from 0.43 at Time 1 to 0.44 at Time 2, this time, however, with a small effect size (d = -0.598).

Table 14. Lexical Density (LD) from Time 1 to Time 2

		LD Time 1 $\bar{X}$ (SD)	LD 2 $\bar{X}$ (SD)	Cohen's d ES
Pen-and-Paper Writing	TA	0.47 (0.02)	0.46 (0.02)	**0.913**
	WL	0.47 (0.04)	0.47 (0.04)	0.381
	WL+TA	0.43 (0.04)	0.43 (0.04)	-0.170
Computer-Mediated Writing	TA	0.43 (0.03)	0.44 (0.03)	**-1.291**
	WL	0.43 (0.01)	0.44 (0.01)	**-0.598**
	WL+TA	0.44 (0.02)	0.45 (0.03)	-0.484

Looking now at lexical sophistication (Table 15), that is, the measure of how many sophisticated lexical items there are in a text -an indicator of text quality-, again minimal differences can be observed. In fact, no significant differences in effect sizes were found from Time 1 to Time 2 for any of the WCF processing groups in both writing environments.

Table 15. Lexical Sophistication (LS2) from Time 1 to Time 2

		LS2 Time 1 $\bar{X}$ (SD)	LS2 Time 2 $\bar{X}$ (SD)	Cohen's d ES
Pen-and-Paper Writing	TA	0.16 (0.04)	0.16 (0.04)	0.143
	WL	0.16 (0.04)	0.17 (0.03)	-0.339
	WL+TA	0.18 (0.02)	0.17 (0.03)	0.363
Computer-Mediated Writing	TA	0.20 (0.03)	0.20 (0.02)	0,221
	WL	0.16 (0.05)	0.16 (0.04)	-0.104
	WL+TA	0.16 (0.04)	0.16 (0.04)	-0.275

The third and final lexical complexity measure was UBER, which refers to the lexical diversity of a text, that is, how many different words appear in it, an indicator of how complex the text is. In line with the results on the other two measures of lexical complexity, very

minimal differences were found between Time 1 and Time 2 for all WCF treatment conditions (Table 16). Nonetheless, the WL+TA groups in both writing environments showed more notable variances. In the P&P writing environment, the scores for UBER decreased from Time 1 (16.22) to Time 2 (15.92) with a large effect size (d = 1.487). In contrast, the scores for UBER of the texts written in the DW environment, increased slightly from Time 1 (16.29) to Time 2 (16.70), with a large effect size (d = -1.201).

Table 16. Lexical Diversity (UBER) from Time 1 to Time 2

		UBER Time 1 $\bar{x}$ (SD)	UBER Time 2 $\bar{x}$ (SD)	Cohen's d ES
Pen-and-Paper Writing	TA	15.80 (0.74)	15.67 (0.88)	0.254
	WL	15.80 (1.28)	15.86 (1.62)	-0.112
	WL+TA	16.22 (1.30)	15.92 (1.38)	**1.487**
Computer-Mediated Writing	TA	16.50 (2.99)	16.62 (2.85)	-0.437
	WL	15.02 (1.32)	15.45 (1.17)	-0.440
	WL+TA	16.29 (1.08)	16.70 (1.32)	**-1.201**

In short, we can observe that WCF processing conditions appeared to play a very minor role on the lexical complexity of the texts written, as minimal differences were found from Time 1 to Time 2. These results were also the same regardless of whether texts were written on paper or on the computer.

- *Syntactic Complexity*

Syntactic complexity was measured in terms of 5 indices, namely, mean length of t-units (MLT), mean length of clauses (MLC), coordinate phrases per clause (CP/C), dependent clauses per clause (DC/C) and complex nominals per clause (CN/C). Results will be discussed by separating each of these components of syntactic complexity, starting first with MLT (mean length of t-units) which constitutes an indicator of the length of production unit, measuring specifically the number of words produced per t-unit.

Table 17. Mean Length of T-Units (MLT) from Time 1 to Time 2

		MLT Time 1 $\bar{X}$ (SD)	MLT Time 2 $\bar{X}$ (SD)	Cohen's d ES
Pen-and-Paper Writing	TA	0.04 (0.01)	0.05 (0.01)	**-0.696**
	WL	0.05 (0.01)	0.05 (0.01)	-0.029
	WL+TA	0.04 (0.01)	0.04 (0.01)	-0.167
Computer-Mediated Writing	TA	0.04 (0.01)	0.04 (0.01)	-0.364
	WL	0.04 (0.01)	0.04 (0.01)	-0.624
	WL+TA	0.04 (0.01)	0.04 (0.01)	0.179

As shown in Table 17, minimal differences were observed between Time 1 and Time 2 for all WCF processing conditions and writing environments. However, in paper-based writing, the participants in the think-aloud only condition improved their MLT score from Time 1 (0.04) to Time 2 (0.05), with a small effect size (d = -0.696).

Table 18. Mean Length of Clauses (MLC) from Time 1 to Time 2

		MLC Time 1 $\bar{X}$ (SD)	MLC Time 2 $\bar{X}$ (SD)	Cohen's d ES
Pen-and-Paper Writing	TA	0.09 (0.02)	0.09 (0.02)	0.415
	WL	0.10 (0.02)	0.09 (0.01)	**0.653**
	WL+TA	0.09 (0.02)	0.10 (0.01)	-0.392
Computer-Mediated Writing	TA	0.08 (0.01)	0.09 (0.01)	-0.297
	WL	0.086 (0.02)	0.089 (0.02)	**-1.003**
	WL+TA	0.08 (0.02)	0.09 (0.02)	-0.511

The second component of syntactic complexity included in the analysis was MLC (mean length of clauses), an indicator of the length of production in terms of number of words per clauses. As shown in Table 18, and in line with the results obtained for the previous measure of syntactic complexity, minimal differences from Time 1 to Time 2 were observed. However,

the WL group experimented notable changes in both writing environments. Specifically, when writing on paper, the texts written by the WL WCF processing group showed a decrease in MLC score from 0.10 (Time 1) to 0.09 (Time 2), with a small effect size (d = 0.653). In contrast, when writing on the computer, the texts written by the WL group showed an increase in the MLC score from Time 1 (0.086) to Time 2 (0.089), with a large effect size (d = -1.003).

Moving on to the third component of syntactic complexity, DC/C (dependent clauses per clauses), an indicator of the ratio of subordination in writing, we can observe in Table 19 how, despite minimal changes across the WCF processing conditions, notable differences can be observed for the think-aloud only group in both writing environments. The texts written by the TA groups both on paper and on the computer showed a decrease in DC/C scores from Time 1 (P&P:0.45, DW:0.52) to Time 2 (P&P:0.43, DW:0.51), with a small effect size (d = 0.723 and d = 0.680, respectively).

Table 19. Dependent Clauses (DC/C) from Time 1 to Time 2

		DC/C Time 1 $\bar{x}$ (SD)	DC/C Time 2 $\bar{x}$ (SD)	Cohen's d ES
Pen-and-Paper Writing	TA	0.45 (0.09)	0.43 (0.09)	**0.723**
	WL	0.41 (0.08)	0.41 (0.10)	0.006
	WL+TA	0.45 (0.13)	0.46 (0.13)	-0.244
Computer-Mediated Writing	TA	0.52 (0.07)	0.51 (0.06)	**0.680**
	WL	0.48 (0.05)	0.48 (0.06)	0.268
	WL+TA	0.50 (0.07)	0.51 (0.07)	-0.408

Concerning the measure for CP/C (coordinate phrases per clause), which is an indicator of coordination in writing, minimal differences can be observed from Time 1 to Time 2 across conditions. However, in the P&P writing environment, the texts written by the WL treatment group experimented a notable decrease in score, with a reduction from Time 1 (0.25) to Time 2 (0.21) and a large effect size (d = 1.105).

Table 20. Coordinate Phrases per Clause (CP/C) from Time 1 to Time 2 for All Treatment Conditions

		CP/C Time 1 $\bar{X}$ (SD)	CP/C Time 2 $\bar{X}$ (SD)	Cohen's d ES
Pen-and-Paper Writing	TA	0.22 (0.08)	0.22 (0.08)	0.060
	WL	0.25 (0.10)	0.21 (0.13)	**1.105**
	WL+TA	0.25 (0.13)	0.24 (0.06)	0.127
Computer-Mediated Writing	TA	0.26 (0.08)	0.27 (0.05)	-0.177
	WL	0.25 (0.10)	0.25 (0.12)	-0.180
	WL+TA	0.26 (0.07)	0.22 (0.07)	0.587

Finally, the last component of syntactic complexity included in the analysis of L2 writing was CN/C (complex nominals per clause), which is an indicator of the relationship between complex nominals and the larger production unit (clauses). Despite minimal changes from Time 1 to Time 2 in general, there are three notable reductions in CN/C scores worth mentioning. First, in the P&P writing environment, the texts written by the WL+TA groups showed a reduction in scores from Time 1 (0.99) to Time 2 (0.91) with a medium effect size (d = 0.721). Along the same lines, the texts written by both the WL and WL+TA groups also experimented reductions in the digital environment. For the WL group, the scores reduced from 1.09 (Time 1) to 0.96 (Time 2) with a high effect size (d = 0.989). Similarly, the WL+TA group also experimented a reduction in scores from Time 1 (1.01) to Time 2 (0.90) again with a large effect size (d = 1.219).

Table 21. Complex Nominals per Clause (CN/C) from Time 1 to Time 2

		CN/C Time 1 $\bar{X}$ (SD)	CN/C Time 2 $\bar{X}$ (SD)	Cohen's d ES
Pen-and-Paper Writing	TA	1.27 (0.45)	1.24 (0.46)	0.428
	WL	1.23 (0.22)	1.14 (0.19)	0.661
	WL+TA	0.99 (0.28)	0.91 (0.16)	**0.721**
	TA	1.04	0.99	0.565

Computer-Mediated Writing		(0.11)	(0.13)	
	WL	1.09 (0.15)	0.96 (0.21)	**0.989**
	WL+TA	1.01 (0.23)	0.90 (0.26)	**1.219**

In short, WCF processing conditions appeared to play a minor role on the syntactic complexity of the texts written, as minimal differences were found in this dimension of syntactic complexity from Time 1 to Time 2. The only notable changes worthy of mention are the slight decreases from Time 1 to Time 2 in terms of syntactic complexity (dependent clauses per clause and complex nominals per clause in this case), an indication of a potential negative effect of WCF processing condition on these dimensions of written production.

V.3. RESEARCH QUESTION THREE (RQ3). EFFECT OF PROCESSING CONDITIONS ON DOP OF WCF ACROSS WRITING ENVIRONMENTS

The third and final research question aimed at exploring the relationship between the type of WCF processing condition and the levels of depth of processing each type induced across writing environments. The results will be reported by looking at the data from both a quantitative (via frequency counts) and qualitative perspective. First, the data from the written languaging tables will be examined, followed by an analysis of the think-aloud transcriptions.

V.3.1. Written Languaging

To recap, two treatment groups involved in the study completed written languaging tables whilst processing the written corrective feedback they received, one of which also completed think-aloud protocols whilst writing. In this section, only the results from the written languaging tables will be discussed, with an analysis of the think-aloud protocols to follow in the next section.

As previously stated in the methodology chapter (IV.3.2.1.), the written languaging tables included four columns in which the participants were required to provide information on the error made in their original text, its corresponding correction, the error code, and an explanation. First, the frequency counts for each column will be discussed, followed by an analysis of the results according to the levels of awareness established in section IV.5.2.1 that is, noticing, reporting, and understanding.

IV.3.1.1. Noticed Errors

Table 22. Frequency of Successful Error Noticing in WL only Processing Condition

	Written Languaging	
	Computer-mediated	**Pen-&-Paper**
$\bar{x}$	75.25	71.83
SD	20.68	32.28

Table 22 shows the frequency of error noticing by the written languaging only group. It can be observed that such level of error noticing (corresponding to the amount of annotated errors in the written languaging tables) was slightly higher in the DW environment (75.25%) as compared to the level of error noticing when writing on paper, in the P&P group (71.83%).

Table 23. Frequency of Successful Error Noticing in WL + TA Processing Conditions

	Written Languaging & Think-Aloud	
	Computer-mediated	**Pen-&-Paper**
$\bar{x}$	80.83	74.50
SD	6.40	17.89

This higher level of noticed errors was also observed in the data corresponding to the written languaging plus think-aloud processing condition, as shown in Table 23. It can be observed that there was a decrease in noticed errors from 80.83% when writing on the computer to 74.50% when writing on paper.

Table 24. Comparison of Successful Error Noticing in WL only and WL +TA Processing Conditions.

	Written Languaging		**Written Languaging & Think-Aloud**	
	Computer-Mediated	**Pen-&-Paper**	**Computer-Mediated**	**Pen-&-Paper**
$\bar{x}$	75.25	71.83	80.83	74.50
SD	20.68	32.28	6.40	17.89

Table 24 shows the comparison of the two treatment conditions and the two writing environments, it can be seen that the written languaging + think-aloud experimental group outperformed the written languaging only group, both in P&P and DW environments, noticing a higher number of errors in their writing (DW= 80.83%; PP= 74.50%).

IV.3.1.2. Error codes

As explained in the Method chapter, participants were asked to code the noticed errors according to type, that is to identify whether errors were grammatical, lexical or if they corresponded to a spelling or punctuation mistake. If a participant believed an error did not fit in to one of these categories, they were asked to note down the error as "other" and elaborate their reasoning.

Table 25 shows that the performance of written languaging only group, was very similar in computer-mediated and in pen-and-paper writing: They successfully categorised 87.98% of noticed errors when writing on the computer and 85.97% when writing on paper. In the case of the written languaging + think-aloud group, the participants in the P&P condition outperformed those in the computer-mediated group, with an average of 84.15% of noticed errors successfully coded, versus 82.53%, in the case of the DW condition. It is worth noting the observed relatively high percentage of success in error coding for both the WL only and the WL+TA groups.

Table 25. Comparison of Successful Error Codes in WL Only and WL+TA Groups

	Written Languaging		**Written Languaging & Think-Aloud**	
	Computer-Mediated	**Pen-&-Paper**	**Computer-Mediated**	**Pen-&-Paper**
$\bar{x}$	87.98	85.97	82.53	84.15
SD	10.77	12.97	19.28	7.36

IV.3.1.3. Explanation of Errors

Participants were required to provide an explanation for the errors they had made in their writing in the last column of the written languaging tables. As shown in Table 26, the

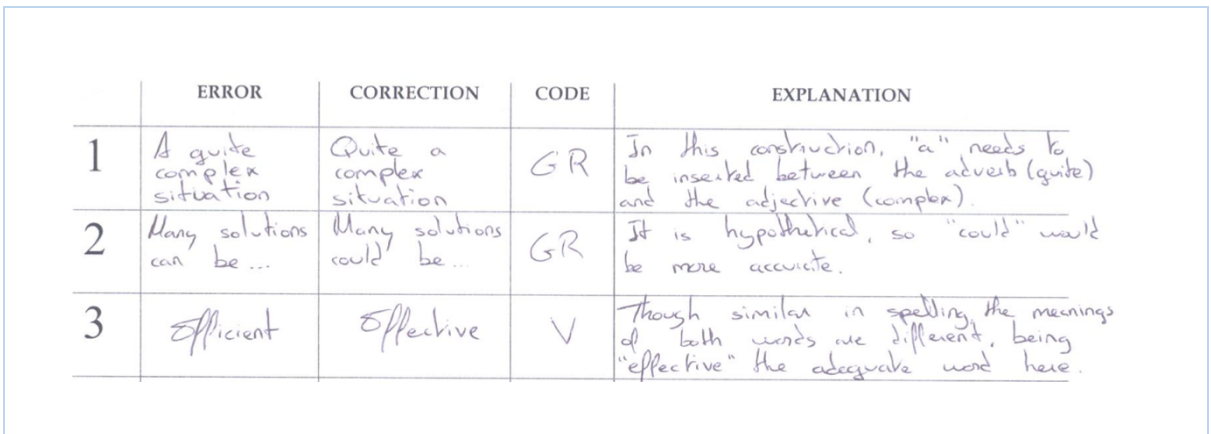
written languaging only groups provided a high number of explanations for their errors, 90.84% in the DW environment 82.11% in the P&P writing group.

In a similar manner, the data of the written languaging plus think-aloud condition also showed a high frequency of error explanations, with participants in the DW environment providing explanations for 85.29% of noticed errors and those in the P&P condition explaining 76.36% of noticed errors.

Table 26 further shows the WL only group were the ones who included most explanations for noticed errors for both the DW condition (90.84%) and the P&P writing condition (82.11%), with the WL+TA group providing slightly less explanations in both the DW (85.29%) and the P&P writing conditions (76.36%). It was also observed that in the P&P environment, the participants in the WL +TA group left explanation column blank more often, as seen in Figures 12 and 13 below. What becomes apparent when comparing these two groups is the high number of blank spaces that appear in the WL tables for the WL+TA group, particularly when concerning the explanation section.

Table 26. Comparison of Explanations Provided for Noticed Errors in the WL Only and WL+TA Processing Conditions

	Written Languaging		**Written Languaging & Think-Aloud**	
	Computer-Mediated	**Pen-&-Paper**	**Computer-Mediated**	**Pen-&-Paper**
$\bar{x}$	90.84	82.11	85.29	76.36
SD	10.87	16.32	15.06	17.18



	ERROR	CORRECTION	CODE	EXPLANATION
1	A quite complex situation	Quite a complex situation	GR	In this construction, "a" needs to be inserted between the adverb (quite) and the adjective (complex).
2	Many solutions can be ...	Many solutions could be ...	GR	It is hypothetical, so "could" would be more accurate.
3	Efficient	Effective	V	Though similar in spelling the meanings of both words are different, being "effective" the adequate word here.

Figure 12. Excerpt taken from the WL only Group (P&P Writing)

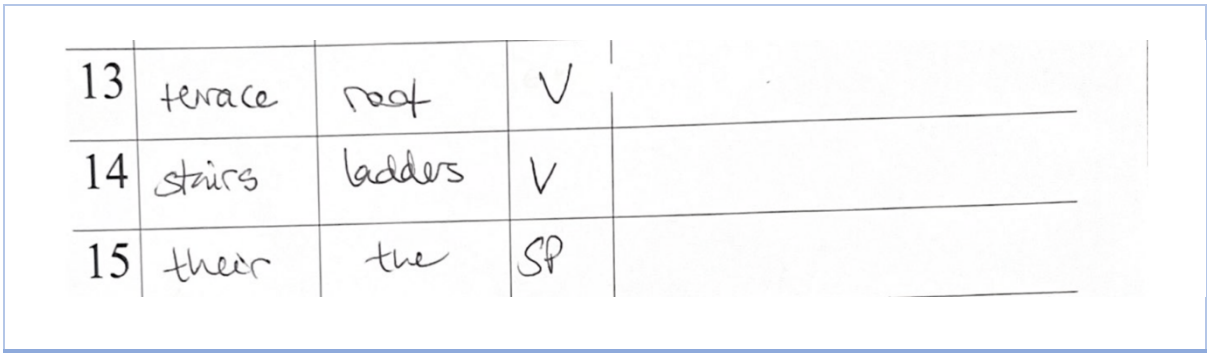


Figure 13. Excerpt Taken from the WL + TA Group (P&P Writing)

IV.3.1.4. Levels of Awareness

As discussed in the Method section (IV.5.2.1), the analysis of the processing of the feedback (as manifested in the written languaging tables) was approached from an awareness perspective on the basis of the coding scheme implemented in Cerezo et al. (2019). Therefore, we categorised the data output from the written languaging tables according to five levels, which corresponded to three global levels of awareness:

- Level 1: Awareness at the level of noticing
- Levels 2 and 3: Awareness at the level of reporting
- Levels 4 and 5: Awareness at the level of understanding

The instances in which the error correction was not noticed were categorised as level 0 (null noticing).

Table 27. Levels of Awareness for WL and WL+TA Conditions

	Written Languaging		**Written Languaging & Think-Aloud**	
	Computer-Mediated $\bar{X}$ (SD)	**Pen-&-Paper** $\bar{X}$ (SD)	**Computer-Mediated** $\bar{X}$ (SD)	**Pen-&-Paper** $\bar{X}$ (SD)
Level 0 (No noticing)	23.13 (21.00)	29.63 (30.45)	19.13 (9.67)	24.06 (16.52)
Level 1	1.39	2.25	0	0.67

	(3.40)	(5.52)	(0)	(1.63)
Level 2	4.96 (8.62)	5.06 (7.96)	11.5 (12.18)	4.80 (4.85)
Level 3	1.39 (3.40)	6.61 (10.65)	0.93 (2.26)	10.17 (6.87)
Level 4	5.20 (5.79)	6.11 (4.47)	5.37 (9.17)	8.44 (6.01)
Level 5	63.93 (24.30)	50.75 (28.47)	63.06 (9.67)	48.75 (16.52)

As shown in Table 27, and solely taking into consideration the data provided in the written languaging tables (the amount of information each participant included), it appears that the DW condition resulted in higher levels of awareness: participants in the DW condition provided a marginally larger quantity of explanations in their WL tables, corresponding to higher levels of awareness (and understanding) according to Cerezo et al.'s (2019) coding scheme.

A fuller analysis of this tendency results from a qualitative perspective. To analyse the data qualitatively, we measured the participants' levels of engagement with the feedback provided (see Table 8, section IV.5.2.3). As the information provided in the tables did not equate to the processing the participants underwent, but to the outcome of the processing (see discussion in the Method section), we analysed the nature of the explanations provided and coded them according to the type of information included in the WL tables.

Reiterating what was previously mentioned in section IV.5.2.1, the explanations for the errors were first categorised according to whether the participants engaged with the error correction or not. If a participant engaged with the feedback provided, the explanations were then coded according to the information provided by the participants and their approach to explaining their errors (e.g., translating, personal reasons, rule formulations, etc.). Therefore, the analysis that follows first looks at whether or not participants engaged with the feedback, and then goes on to explore those instances in which engagement with the corrections was present, by qualitatively analysing their explanations and providing examples.

Table 28. Types of Engagement with WCF in the WL Tables

		Written Languaging		Written Languaging & Think-Aloud	
		DW $\bar{X}$ (SD)	**P&P** $\bar{X}$ P6P (SD)	**DW** $\bar{X}$ (SD)	**P&P** $\bar{X}$ (SD)
	Evidence in WL table				
Participant does not engage with the EC					
	1. Leaves a blank space	23.13 (21.00)	32.10 (29.98)	18.20 (10.79)	25.68 (16.05)
	2. Annotates the error and correction with no further analysis	-	-	-	6.25 (10.23)
Participant engages with the EC					
Disagrees with the EC	1. "I disagree" in the explanation section, but no further information provided.	-	0.45 (1.10)	-	3.22 (4.99)
	2. Provides a written explanation for their disagreement	-	1.80 (4.41)	2.56 (6.27)	0.57 (5.17)
Agrees with the error correction	1. Explanation section left blank or evidence of not being able to provide an explanation	1.38 (3.40)	10.34 (12.98)	3.18 (4.93)	3.35 (5.17)
	2. Evidence in the explanation section that a new L2 form has been accepted. "I didn't know that word/form"/ "I	14.72 (17.53)	1.66 (2.58)	4.45 (7.20)	6.85 (6.74)

	didn't know how to say it"				
	3. Translation included in the explanation section	6.76 (11.71)	1.11 (2.72)	2.5 (6.13)	2.14 (3.90)
	4. Personal reasons (i.e., rushing, always make this mistake.)	6.74 (7.70)	0.90 (2.20)	31.92 (36.25)	3.22 (6.44)
	5. Rule explanation/Formulation included in the explanation section	47.25 (34.89)	53.34 (29.88)	37.17 (37.32)	47.06 (18.55)

Table 28 shows the frequencies for the instances in which no engagement with the feedback were found, as manifested in blank spaces in the WL table. The P&P groups in both treatment conditions were the ones who left out more error corrections (WL=32.10%; WL+TA=25.68%). In turn, with slightly lower frequencies, the participants in the DW condition excluded an average of 23.13 % of errors in the WL group and a lower 18.20% of errors in the WL + TA group. Worthy of note are the overall lower frequencies in the WL + TA for each writing environment, suggesting that processing feedback while thinking aloud played a role in the noticing of and engagement with error corrections.

Concerning the types of engagement found in the written languaging tables, worthy of note are the disagreements expressed by participants with the error corrections provided, mainly found in P&P writing. In these cases, participants disagreed with the corrections but did not provide reasons why (WL=0.45%; WL+TA= 3.22%). In the cases in which participants disagreed but provided a reason justifying their disagreement, there were also several DW participants in the WL + TA group who demonstrated resistance towards the feedback and provided reasons for their disagreement with the error corrections provided (2.56%). Although the frequencies for these disagreements are relatively low, these findings were rather unexpected, and some examples can be found below. As shown in Figure 14, the participant shows two signs of disagreement with the corrections provided. In the first case, the participant believes that the use of "could" is related to a possibility and, therefore, "will" was more appropriate, as the participant explained they were referring to an action s/he was going to do for certain (not something that was a possibility). In the second example, the participant

believes it is better to use the term “take” instead of “free”, which was the error correction provided.

In Figure 15, in the first example, the participant explicitly states disagreement with the correction, which related to the misuse of lower-case letters, explaining that the word had in fact been correctly written with a capital letter. In the second example, the same participant states that s/he is unaware of why the preposition is incorrect and thinks it is “OK”. Finally, In Figure 16, in the first example, the participant believes that both words are acceptable and in the second example, the participant explains that s/he used “do it there” so as not to repeat the same phrase they had used in the first line.

	Error	Correction	Code	Explanation
1	Will	Could	Grammar	I was going to do it, like there wasn’t a possibility.
2	Take	Free	Grammar	I think it was better to write take instead of free, I usually use take more than free

Figure 14. Examples of Disagreements from a Participant in the DW Environment (WL+TA)

	Error	Correction	Code	Explanation
1	For	for	SP	I disagree. I have written correctly the word after the full stop
2	Until	To	G	For me it is okey, I don’t know why it is wrong

Figure 15. Examples of Disagreements from a Participant in the P&P Environment (WL+TA)

	Error	Correction	Code	Explanation
1	One	Person	V	I think both words fit
2	Do it there	Put out the fire there	V	I used "do it there" to not repeat the same situation in the first line

Figure 16. Examples of Disagreements from a Participant in the P&P Environment (WL only group)

Table 28 shows the instances in which participants engaged with the feedback provided by agreeing with the correction provided. The first type of this engagement coded corresponds to when participants left the explanation column blank or included some evidence of not being able to explain the correction. In most cases, this outcome corresponded to participants explicitly claiming they were not sure how to explain the correction provided (see Figure 17). For this category, the participants in the WL only group in the P&P writing condition were the ones who had the highest frequency (10.34 % of errors) of which some examples are provided below.

	Error	Correction	Code	Explanation
1	Grow bigger	Grow any bigger	V	I do not really know why "any" is necessary
2	Had	Has	GR	Not sure why this one is wrong

Figure 17. Examples of Participants not Being Able to Explain the Error Corrections (P&P Writing, WL Only Group)

The second type of engagement corresponded to acceptance of a new L2 form. In the WL group, it was the DW condition who had a higher frequency for this type of engagement (14.72 % of the errors). The notes concerning an indication of a new L2 form mainly corresponded to when participants mentioned not previously knowing the word/form, or not knowing how to say in in their L2 (Figure 18).

	Error	Correction	Code	Explanation
1	Are open air	Have fresh air	GR	Didn't know the correct collocation
2	Reach to the top	Reach the top	GR	Didn't know the correct collocation

Figure 18. Examples of "new" Knowledge from a Participant in the DW Environment (WL Only)

The third type of engagement concerned references to the participants' L1 (in this case Spanish) or translations between their L1 and L2. Participants in all feedback processing conditions included translations in their written languaging tables or mentions of their L1 in regard to the error correction provided, with the DW, WL group showing the highest frequency of this type of explanation (6.76% of errors). In Figure 19, example 1, we can see how the participant refers to the word "resting" as being more "Spanish", and in the second example, the participant mentions the construction being used in informal Spanish.

	Error	Correction	Code	Explanation
1	Resting	Remaining	L	Resting seems to be a more Spanish sense, since resting is a non-finite verb form of "to rest"
2	The most safest	The safest code	GR	A redundant construction, normally used in informal Spanish

Figure 19. Examples of Participants Referring to their L1 from a Participant in the DW Environment (WL Only)

One of the most striking results corresponded to the fourth category in Table 28, in which participants provided personal reasons for their errors, rather than the requested metalinguistic comments. A trend in the data revealed that the participants writing on the computer appeared more prone to justifying their errors by providing personal explanations (6.75% of errors in the WL group and a large 31.92% of errors in the WL+TA group). They

mostly mentioned that they were writing too fast thus resulting in the error or that they commonly made the same mistake. Below (in Figures 20 and 21) are some examples taken from DW group in which some of the personal reasons provided can be seen. In Figure 20, the participant mentions writing too fast and making the error often. In Figure 21, the participant states that the error was made "in the moment" and mentions writing in a hurry as the reason why. In example 3 of Figure 21, the participant states that the error was made due to confusing the term with the Scottish dialect, as they had been listening to this accent prior to completing the writing task.

	Error	Correction	Code	Explanation
1	His	Him	Grammar	I wrote too fast
2	In	On	Grammar	I forgot to write it because I wrote too fast. Common error

Figure 20. Examples of "Personal Reasons" Taken from a Participant in the DW Environment (WL+TA)

	Error	Correction	Code	Explanation
1	Ran	Run	Spelling	This was just a mistake made in that moment
2	In the	On the	Grammatical	I was in a hurry to write this task, so the mistake was in the moment
3	Old one	Man/Woman	Vocabulary	Confused by the Scottish dialect, which I was listening to before completing the task

Figure 21. Examples of "Personal Reasons" Taken from a Participant in the DW Environment (WL+TA)

The fifth type of engagement found (the final column in Table 28), corresponds to engagement as manifested by rule-formulation (at times successful, at times not), in which participants provided explanations which were much more metalinguistic in nature. Below are some examples showing the range of rules the participants provided. Data from all feedback treatment conditions showed a high frequency of this type of explanations, with the participants in the WL only group including a higher number of rule-related explanations (47.25% for the DW condition and 53.34% for the P&P condition). Despite the fact that the WL+TA condition showed slightly lower frequencies (DW: 37.17% and P&P: 47.06%), it is important to remember that this group also engaged with the error corrections via think-aloud protocols and thus, the following section (V.3.2.) will report the TA data and analyse how this influenced the information incorporated into the WL tables.

	Error	Correction	Code	Explanation
1	Probably, I would	I would probably	GR	Wrong word order. Adverbs are used between the auxiliary verb and the main verb.
2	There is not a fire extinguisher	There is no fire extinguisher	GR	I have used a determiner for an uncountable noun

Figure 22. Examples of "Rule-Formation" taken from a Participant in the P&P Writing Environment (WL+TA)

	Error	Correction	Code	Explanation
1	Must	Should	GR	It is not an obligation but an advice
2	Would go down	Am going	GR	After "while", a continuous tense is placed

Figure 23. Examples of "Rule-Formation" Taken from a Participant in the P&P Writing Environment (WL Only)

As shown in Figure 24, the participant includes a rule on the use of "fog" and "smoke" (Example 1) and also discusses the incorrect use of the comparative form "less" as opposed to the correct superlative form "least".

	Error	Correction	Code	Explanation
1	Fog	Smoke	V	"Fog" is a meteorological phenomenon, "smoke" is the correct word
2	Less	Least	GR	"Less" is the comparative form, but in this context, it should be "least", the superlative one

Figure 24. Examples of "Rule-Formulation" Taken from a Participant in the P&P Writing Environment (WL Only)

In Figure 25, the participant successfully discusses two phrasal verbs, explaining the differences in their meaning and in terms of the contexts in which each one should be correctly used.

	Error	Correction	Code	Explanation
1	Turn off the fire	Put out the fire	G	The correct phrasal verb here would be "put out" since "turn off" is to refer to turn off "a light", "television", etc. We use "put out" to refer to fire
2	Taken on	Taken from	G	We use "take from" when we take something

				from somewhere, not something that is on something (on a table); it is referring to the fire-extinguisher that we take from the first floor

Figure 25. Examples of "Rule-Formulation" Taken from a Participant in the DW Environment (WL + TA)

Finally, in Figure 26 (example 1), the participant discusses the use of "will" and "would" and decides that "would" is more appropriate as the sentence was a hypothetical situation. In example 2, the participant admits that the word order of the sentence was incorrect, and the word "again" needed to go after the object.

	Error	Correction	Code	Explanation
1	I will do	I would do	GR	"Would" because it is a hypothetical or fictitious situation
2	I use again the fire truck	I would use the fire truck again	GR	The word order of the sentence is wrong, "again" must go after the object.

Figure 26. Examples of "Rule-Formulation" Taken from a Participant in DW Environment (WL Only)

V.3.2. THINK-ALOUD PROTOCOLS

In order to analyse the feedback processing in the think aloud condition, the data was transcribed and coded according to levels of depth of processing (coding scheme can be found in Table 7, in IV.5.2.2 of the Method chapter. The results will first be discussed quantitatively

(frequency counts of levels of depth of processing- high, medium, low, and null-) and then qualitatively in terms of language-related episodes in the data.

Table 29. Levels of DoP for Noticed Errors in the TA Transcriptions

	Think-Aloud		**Written Languaging & Think-Aloud**	
	Computer-Mediated $\bar{X}$ (SD)	**Pen & Paper** $\bar{X}$ (SD)	**Computer-Mediated** $\bar{X}$ (SD)	**Pen & Paper** $\bar{X}$ (SD)
Null	17.41 (8.33)	8.79 (12.53)	0 (0.00)	6.35 (15.55)
Low	48.61 (21.19)	57.42 (7.76)	23.90 (19.60)	30.37 (19.54)
Medium	8.68 (9.49)	22.07 (12.96)	31.65 (30.11)	20.55 (16.50)
High	25.30 (14.20)	11.72 (13.08)	44.45 (41.31)	41.91 (19.04)

V.3.2.1. High Levels of Processing

Table 29 shows that the WL+TA group were the ones who had the highest frequency count of high levels of processing both for the DW and P&P writing environments (44.45% and 41.91% respectively). Much lower percentages of high levels of processing were observed in the TA only group (25.30% for the DW and 11.72% for P&P writing). This higher engagement, seemingly facilitated by the availability of the written languaging table, was manifested in the participants' LREs, which showed instances of:

- successful formulation of rules;
- hypothesising about the error corrections;
- successfully translating the error corrections and providing a correct metalinguistic explanation;

- demonstrating high levels of cognitive effort, as manifested in the time spent engaging with the error corrections and providing correct metalinguistic explanations;
- disagreeing with the correction given but providing a correct metalinguistic explanation to justify their disagreement.

As an example of these instances of high levels of processing, Figure 27 contains data from a participant from the WL + TA group in the P&P writing environment where s/he discusses an error related to subject omission in the sentence. In the transcription, the participant mentions forgetting to include the subject for the verb "could". The participant then moves on to explain such omission by relating this error to L1 influence as in Spanish the subject can be omitted because the form the verb takes makes it clear who is performing the action, although s/he is aware that such omission is not possible in English. This explanation is taken as evidence of cognitive effort and metalinguistic reasoning, hence its classification as an instance of high DoP. Worthy of mention is the affordances of the TA data and the stark difference between the information included in the written languaging table compared to the TA transcription. Without the availability of the TA transcription, it would be impossible to truly measure the depth of processing of the participant's engagement with the feedback provided, as only the outcome of the feedback processing is manifested in the WL table (in the explanation column). As previously mentioned, the TAs were crucial for measuring DoP levels and shedding light on the ways in which participants approached feedback processing and engaged with the error correction provided.

	Error	Correction	Code	Explanation
	Subject	He/she could	GR	Influence of Spanish
TA Transcription In the following one…and last one (2) I…I omitted the subject and…I only wrote the verb…the main verb (2) and…maybe…it's because (3) I wrote a subject that…the main subject before…and…I don't know…I completely forget…forgot the…the subject again (4) maybe its…maybe it's because…because of the literal…because of the influence of Spanish because we don't need to…to write the subject…hmm…so many times because it's (3) hmm (2) the subject is (3) it's know when you're speaking so.				

Figure 27. Example of High DoP in P&P Writing Environment (WL + TA)

In the second example of High DoP (Figure 28), again taken from P&P writing, although this time in the think-aloud only condition, it can be observed how the participant discusses the erroneous use of the word "ancient" to describe a person. The participant states that having been given the opportunity to think about it, perhaps this term is more appropriate when describing "buildings or things like that", as opposed to people, and agrees that the term "old" is therefore more adequate. The opportunity to process the error allowed the participant to reflect on the uses of the word "ancient" and compare it to the provided correction "oldest". This led the participant to formulate a hypothesis about when each word should be used.

TA Transcription: original	**TA Transcription: Translation**
The oldest person (2) y yo he puesto ancient (2) creo que es de vocabulario si…porque ancient ahora que lo pienso creo que es mas para edificios o como cosas así…no para personas entonces para persona sería old…no ancient…vale sí…es verdad…no sé por qué he puesto esto (2) the ancient person	The oldest person (2) and I have put ancient (2) I think this is vocabulary yes...because ancient now that I think about it I think that it is more for buildings or things like that...not for people so for a person it would be old...not ancient...ok yes...it's true...I don't know why I put that (2) the ancient person

Figure 28. Example of High DoP in P&P Writing Environment (TA Only Group)

An example taken from a participant in the WL + TA group in the DW condition appears in Figure 29. In the WL table, the participant explains the error as "confusion of uses of to-infinitives, also a mistake in the moment" but produces a more complete explanation in the TA protocol. The transcription of the TA data shows that the participant initially laughs upon seeing the error correction, an indication that s/he was aware of the error made and likely knew the explanation behind the correction. The participant discusses not being used to using infinitives after a colon in his/her writing but wanted to include a list of two options. The participant states that s/he is sometimes confused by the use of 'to infinitives' and 'bare infinitives' but also that upon seeing the correction, s/he was fully aware that the bare infinitive was necessary in this particular sentence (note that no mention of bare infinitives was included in the WL table). In this example, it is clear again that the TA transcription is the instrument

that provides the most comprehensive view of DoP and that thanks to the TA protocol it is possible to discern how the participant came to the explanation included in the WL table. This instance was coded as high DoP as the participant engaged with the error correction for an extended amount of time, particularly when compared to other TA protocols. In addition, the participant provides metalinguistic information when discussing the use of a colon and infinitives.

	Error	Correction	Code	Explanation
	To go	Go	GR	Confusion of uses of to-infinitives, also a mistake in the moment

TA Transcription

Hmm…then…the next one…hmm (2) yeah…like…I know why [LAUGHS]…it's just…as I said hmm (3) no…I…ohh (2) there's a mistake I didn't…write (2) ok…so I need to…ok so I am gonna write it here (7) so (4) I'm gonna…to go… go (3) ok…so…hmm (2) the…the sentence is then I would ask that person to make a choice…to…I…and I wrote…to row…to go…downstairs and get out of the building safely…so hmm…here…hmm…the correction is that it is without the to…hmm…the to…the particle to…so it's go…downstairs or…ok…so hmm…normally in my head…when I…think about infinitives I…think of infinitives with the to…to infinitives…and…I think…I'm not…I am not very used to writing an infinitive after a colon (2) I…here I wrote…a…a colon because I…I wanted to…to imply that there were two…two options…for that person to make…so it was this…this person has a choice…this…or…this…maybe…if I would have…If I had written like…hmm…if I had written…either this…or this (2) I would…I would have seen it clearer in my mind but hmm…but at that moment… it's like to go…or to come [LAUGHS] you know? Hmm…I think also like…a moment…a moment of confusion…mistake in the moment hmm (2) because when I see…see it written…by another person…I see why and I…and I know that…it's like that…but at that specific moment when…I…did this….I wrote this…text (2) I [LAUGHS] I just…saw it…very clearly that it was…like this so…I…think…this is…hmm…grammatical (3) [writes in the WL table] and…hmm (2) I am gonna write…confusion…confusion of uses of to infinitives…in…fin…itives (2) and…bare…infinitives (2) because…if it is true that

sometimes I…I am confused about the usa…the uses…oh my god…and…also…hmm…also a mistake… [writes in the WL table] in the moment (3)

Figure 29. Example of High DoP in DW Environment (WL + TA)

In the next example (Figure 30) taken from the TA only, DW data, the participant discusses the incorrect use of the word "ask" before the direct object "closing". The error has been corrected to replace "ask" with "suggest", thus leaving the -ING form of the gerund intact. Upon viewing the error correction, the participant believes the word has been changed as it is more formal, but also states that the verb "suggest" requires the use of an -ING form and therefore understands why "suggest" is more adequate, successfully formulating the rule.

It is worth adding that very few examples of High DoP were found in the data from the TA only groups, particularly in the DW data, and the TA transcriptions in general were of a much shorter length when compared to those from the WL + TA groups, thus suggesting, as previously mentioned, that the WL table served as a crucial guide in directing and engaging participants to process each error correction more deeply.

TA Transcription: Original

And the fumes are already out of the building…I would not…su-ggest…or ask... (3) maybe suggest is like more…formal? (3) hmm (3) I would not suggest…closing any doors…yeah…I think this is clear and…maybe it's…another way to say it…like…it's clear that after suggest…goes…[LAUGHS]…hmm…a verb with ING so (2) I think this is another way to say it (5)

Figure 30. Example of High DoP in DW Environment (TA Only Group)

V.3.2.2. Medium Levels of Processing

The data from the P&P writing condition, showed similar frequency counts of medium levels of DoP, with the TA only group having medium DoP for 22.07% of their noticed errors, and the WL + TA groups processing 20.55% of noticed errors at a medium DoP level. In contrast, the DW data differed considerably, with the TA only group showing medium DoP for

only 8.68% of noticed errors and the WL + TA group processing 31.65% of their noticed errors at a medium level. It should be recalled that medium DoP corresponded to participants:

iv. translating the error correction (successfully or unsuccessfully) but not providing a metalinguistic explanation or any further information.
v. discussing having made the same mistake previously in the text (repeated error) and providing a correct explanation.
vi. attempting to explain the error correction metalinguistically but in terms of cognitive effort, very little time was spent engaging with the target item and/or the attempt to explain was abandoned.

Some examples of medium DoP are provided below. In the first example (Figure 31), taken from the P&P writing, WL + TA group, the participant discusses having made the same mistake previously in the text and provides a short explanation regarding the necessary use of the pronoun "this" instead of "it". What is particularly interesting in this example and, shown in Figure 31, is that the participant did not include this error in the WL table, as manifested in the blank space left in all columns. As mentioned in earlier sections, the WL+TA group left many blank spaces in their WL tables, despite the participants reflections on the errors in the TA protocols. Therefore, the availability of the TA transcriptions made it possible to view the errors that had been processed but not included in the WL tables.

Error	Correction	Code	Explanation
-	-	-	-

TA Transcription: Original	**TA Transcription: Translation**
El siguiente error es similar al que ya he cometido en el…hmm… [looks back on the first sheet (4)] en el numero 4…hmm…yo pongo but it…pero…hmm (3) la opción co…correcta es this…porque me estoy refiriendo a lo que había…hmm…mencionado anteriormente en el texto (7)	The next error is similar to the one I have already made in the...hmm... [looks back on the first sheet (4)] in number 4...hmm...I put but it...but...hmm (3) the correct option is this...because I am referring to what was...hmm...mentioned before in the text (7)

Figure 31. Example of Medium DoP in P&P Writing Environments (WL + TA)

The second example (Figure 32), also taken from the P&P data, this time from the TA only group, shows the participant discussing a vocabulary error. In this case, the participant translated the word into Spanish and stated that, at the time, s/he was not aware of how to say it in English and so, s/he wrote it in the L1, with no further information provided.

TA Transcription: Original	**TA Transcription: Translation**
Aquí tengo un error de vocabulario porque no sabia como poner extinctor y en el momento lo puse tal cual en español [LAUGHS] y me lo corrige como es en inglés pero bueno (6)	Here I have a vocabulary mistake because I didn't know how to say extinguisher (translates into Spanish) and in the moment I just put it like that, in Spanish [LAUGHS] and they have corrected it in English but yeah (6)

Figure 32. Example of Medium DoP in P&P Writing Environment (TA Only)

Moving on now to the DW data, the first example (Figure 33) is taken from the WL + TA condition and shows a participant processing an error related to the incorrect use of prepositions. The participant mentions having made this error before in the text and having written "in" instead of "on". The error is correctly identified as a grammatical one but further information on why "on" is more appropriate than "in" for this particular sentence is not provided. In addition to mentioning having already made this error, the participant also provides a reason for making the error, saying that perhaps s/he was writing too fast and therefore forgot to write "on" instead of "in". As previously mentioned, a common finding in the data, particularly in the DW data, was the provision by the participants of more personal (rather than purely linguistic) reasons for their errors, attributing them to causes such as not paying enough attention or writing too quickly, an example of which can be seen in Figure 33.

	Error	Correction	Code	Explanation
	In	On	GR	I forgot to write it because I wrote too fast. Common error.
TA Transcription: Original				

The fifth (2) on the second...floor…yep…I wrote in instead of on (3) [writes in the WL table] it is grammar (2) I am sure (2) and the explanation is…maybe there…maybe it's the same as the…as the…the previous one…I wrote it too fast and maybe I (2) I forgot and…I…am going to write…too…that it is…it is a common error (12) hmm (2)

Figure 33. Example of Medium DoP in DW Environment (WL + TA)

Finally, another example taken from the DW data, in this case from the TA only group, can be seen in Figure 34. In this LRE, the participant discusses the incorrect use of the pronoun "his" instead of "their". The participant reads the error and the correction, agrees ("yes...sure sure sure") and provides an extremely brief explanation of why "their" is more appropriate. In this case, because "they are two". However, rather than elaborating more in terms of a metalinguistic explanation, the participant explains that perhaps s/he was not thinking. Again, in this example, and in line with the trend observed in the DW data, the participant opts for providing a more personal reason for the error, rather than explaining why "their" is more adequate than the use of "his".

TA Transcription: Original

Anyway [SIGHS]…having saved this person…I would take both children…and having asked…his mother…or their mother…yes…sure sure sure [NODDING]…their mother…cause they are two…I dunno…I wasn't thinking maybe (2)

Figure 34. Example of Medium DoP in DW Environment (TA only)

V.3.2.3. Low Levels of Processing

The TA only group showed the highest percentage of low levels of processing both for P&P (57.42%) and DW writing (48.61%). Slightly lower frequency counts were found for the WL + TA group in which the P&P data showed low levels of processing for 30.37% of noticed errors and the DW data showed low levels for a smaller 23.90% of noticed errors. Reiterating what was previously stated in the method section, low levels of DoP were characterised by participants:

iv. reading or repeating the error correction;
v. simply agreeing (or disagreeing) with the correction provided and not providing any further information;
vi. not understanding the correction provided (as manifested by signs or verbalisations).

In Figure 35, an example taken from the P&P writing condition (WL + TA group) shows that the participant reads the error and the corresponding correction, recognizes that it is a grammar error and then simply agrees with the correction, without providing any further information as to why one term is more appropriate than the other. It is worth noting the contrast in the information provided in the WL table and the TA transcription. Thus, the participant mentions the need of the superlative form in the WL table, yet no mention of this is included in the TA protocol, in fact, the participant makes no metalinguistic reference at all and simply agrees with the correction before moving on to the next error. Thus, in this particular instance, the TA transcription alone has been classified as low DoP due to the lack of metalinguistic explanation and engagement with the error correction. However, with the complementary information provided by the WL table, it is clear that, although the participant did not verbalise the need for a superlative form, s/he had clearly thought about this and therefore included the information in the WL table. Interestingly, this contrasts the example provided in figure 31 for example, in which the TA protocol is the introspective measure that provides essential information the WL does not. In this case (figure 35), the WL table provides more details than the TA verbalisation and even shows that the student did reach a metalinguistic conclusion as to why the error correction was necessary,

Error	Correction	Code	Explanation
Safer	Safest	GR	I agree. The superlative is needed
TA Transcription: Original		**TA Transcription: Translation**	
Hmm…el siguiente (5) me he perdido…ah vale…por aquí [LAUGHS] I would send the fire-fighters to climb to the second floor as it is the safe…he escrito (3) safer [writing as		Hmm...the next (5) I'm lost...ah ok...here [LAUGHS] I would send the fire-fighters to climb to the second floor as it is the safe...I wrote (3) safer [writing as she speaks]	

she speaks] hmm…en vez de … [COUGHS]…safest…es un…un grammar mistake…I agree [writes down the information in the WL table (2)] hmm (20)	hmm... instead of [COUGHS]...safest...it is a...grammar mistake...I agree [writes down the information in the WL table (2)] hmm (20)

Figure 35. Example of Low DoP in P&P Writing Environment (WL + TA)

In the next example shown below [7], taken again from the P&P data in the TA only condition, the participant spends a very minimal amount of time processing the error correction and simply reads and repeats the correction provided, without any further explanations regarding the error type or the reasons for making this error. This was a common trend in the TA only groups, in which the time spent on engaging with the feedback was much shorter as compared to the WL + TA groups, proving again the crucial role of the WL table in enhancing the feedback processing stage.

[7]

TA Transcription: Original
Vale puse on the roof y lo habéis cambiado por roof terrace…the roof terrace (3)

The following example (Figure 36), taken from the DW data (WL + TA group), shows the participant discussing a preposition error, which has incorrectly been categorized as a lexical error (as opposed to a grammatical one). In this example, the participant shows signs of doubt when processing the error correction (as manifested in the repetition of "hmm" throughout the LRE). The participant read the error correction ("of it") and simply stated that the error is a preposition error, but no further analysis nor explanation is provided. The signs of doubt from the participant and the incorrect coding of the error suggest that s/he did not understand why the error has been corrected in this way and, despite not explicitly saying so, the verbalization indicates this lack of understanding.

	Error	Correction	Code	Explanation
	Of it	Delete	L	The error could be an error in

	the use of the preposition
TA Transcription: Original Ok code…so…ok…I think this one…maybe…it is…of it…could be…hmm (16) hmm (3) hmm (7) explanation…hmm (3) [writing in the WL table] the error (2) could be (5) hmm…an error in…hmm (12) error in…the…hmm…the…hmm (3) preposition (3)	

Figure 36. Example of Low DoP in DW Environments (WL + TA)

The next example [8], taken again from the DW data but this time from the TA only group, shows a participant reading an error related to the incorrect use of a preposition. The participant had originally written who is "in" this floor, rather than who is "on" this floor. However, by simply looking at the TA transcription, there is no evidence of the error the participant had made, and s/he simply reads the correction and moves on to the next error, without providing any further analysis nor explanation. Again, this example demonstrates the lower level of engagement as manifested in terms of time spent on task for the TA only group, especially when compared to the longer LREs found in the WL + TA groups data.

[8]

TA Transcription: Original who is also…on…this floor (3)

V.3.2.4. Null Levels of Processing

The final category of levels of DoP corresponded to the null level, that is, instances which did not fit into the criteria established for high, medium, and low DoP. These verbalizations (or lack thereof) corresponded to instances in which participants showed signs of ignoring the errors and, rather than stopping and discussing the corrections, they moved on without providing any kind of explanation. Although this happened rather infrequently, in the data by the TA only groups this was more common, as 17.41% of errors were coded as null in the DW condition and 8.79% in the P&P writing condition. For the WL + TA groups, only 6.35% of the errors were coded as null for the P&P writing condition, and no instances found for the DW condition.

In Figure 37, an example is shown from the WL + TA group, P&P writing condition in which the participant has included an error correction in the WL table but has not discussed it in the TA protocol, nor has s/he provided an explanation for the error in the last column of the WL table. The lack of LRE in the TA protocol makes it difficult to establish why the participant decided not to discuss this error. The participant repeated this tendency on a number of occasions and was the participant who had the most "black spaces" both in the explanation section of the WL table and the TA transcriptions.

	Error	Correction	Code	Explanation
	Breath	Breathe	SP	
TA Transcription: Original				

Figure 37. Example of Null DoP in P&P Writing Environment (WL + TA)

The following examples [9] and [10] taken again from the TA only group in the P&P writing environment, shows the participant has generalised the errors made by categories and, rather than discussing them one by one, as advised to, the participant simply comments on the fact that many of the same errors have been made repeatedly (prepositions in this particular case). These instances were categorised as null due to the fact that the participant did not discuss each error individually, nor did s/he attempt to provide an explanation for each error made. This tendency to group errors by categories was commonly found in the data by the TA only groups, where, rather than commenting on errors individually, one by one, they opted for making more general comments. In this particular example, the error corrections were not subsequently incorporated into the rewriting, interpreted as lack of processing and consequent internalization of forms this approach led to.

[9]

TA Transcription: Original
Preposiciones…sé que son…todos…la mayoría de los errores que tengo son de preposiciones y es por eso…por lo que he dicho (2)

The last example comes from the DW condition and was taken from the TA only group, as the WL + TA showed no signs of null DoP. In [10], the participant mentions number of

errors made (four in total) but does not stop to discuss them individually, nor does s/he read the corrections provided. The participant simply says s/he "didn't realise" and proceeds to end the TA protocol. Only one out of the four errors ignored was not incorporated into the rewriting, with the three remaining errors successfully corrected. Again, in this example, we can see how the TA only group tended to group the errors together, rather than discuss them individually, with a lower level of processing observed in the data when compared to the data provided by participants in the WL + TA groups, who tended to discuss each error correction individually.

[10]

TA Transcription: Original
The rest of mistakes I…I…didn't realise (3) that's it.

Summarising the results related to Research Question 3, some of the main findings when looking into the effects WCF processing via TAs and WL tables were:

- The introspective measure which included a combination of WL and TA was the most successful in leading to higher deeper levels of processing.
- The TA protocols used in isolation were not as effective in engaging the participants in WCF processing.
- The WL tables used in isolation engaged participants in WCF processing, but the WL data failed to shed light, from a research point of view, on the levels of depth of processing.

In terms of modality, differences were found between DW and P&P WCF processing, with the DW group providing more personal reasons for their errors and the P&P group providing more metalinguistic explanations.

CHAPTER VI. DISCUSSION

This chapter presents a discussion of the results obtained according to the three research questions guiding the study. To recapitulate, research question 1 attempted to shed light on the effects of composing medium (pen-and-paper versus computer-mediated environments) on the complexity, accuracy, and fluency of the texts written by the participants. Research questions 2 and 3 focused on the effects of three feedback processing conditions on writing products (RQ.2) and on the levels of depth of processing (RQ.3), with a secondary methodological aim exploring the affordances of the two introspective measures implemented in the study (namely, written languaging tables and think-aloud protocols) individually and combined.

VI.1. RESEARCH QUESTION 1 (RQ1). EFFECT OF WRITING ENVIRONMENTS ON L2 TEXT CHARACTERISTICS

The first research question addressed potential effects of writing environments (writing in a traditional pen-and-paper versus a computer-mediated writing environment) on undergraduate student's L2 written production. It was found that the texts written in the screen-based condition, showed significantly higher initial accuracy, in line with results found in a study by Vasylets et al., (2022) in which participants writing in the DW environment produced almost half as many errors as those in the P&P environment. Similar results were also found in a study by González-Cruz et al., (2022), in which differences in the overall percentage of global errors were observed, with the texts produced in P&P writing showing a significantly higher number of global errors than those in the DW environment. Yet, texts written in both environments were very similar in terms of number of words and time spent on task.

These findings can be attributed to two principal advantages that DW presents, crucially including the participants' access to a number of writing tools embedded within *GoogleDocs* (the application which was used for the current study). Thus, in terms of mechanical errors, such as spelling or punctuation, a significant reduction was observed in the texts written on the computer as the result of the availability of the spellcheck function of the writing app. Additionally, a DW environment allows students to easily self-edit their texts, going back over the text written, to revise and rewrite. This is facilitated by the fluidity this type of writing condition bestows and the ability it offers to delete words or fragments of texts, enabling writers to have a clean draft of their writing at all times. However, despite the results of this study and

those of previous research (González-Cruz et al., 2022, Vasylets et al., 2022) pointing in the same direction, the findings should not be interpreted as proof of the existence of a direct link between digital writing and language learning. That is, although DW conditions seem to result in more fluent and accurate writing (less errors are made thanks to spelling and grammar checks), more research needs to be carried out to ascertain whether or not these benefits withstand in the long-term, and whether they also apply to using writing as a means of language learning. As many of these errors are corrected automatically, learners' attention may not be fully drawn to their language-related problems and many issues may go unnoticed. Thus, research may benefit in exploring L2 learners' awareness of digital writing tools, such as *spell check*, in order to explore to what extent these types of instruments may lead to language learning.

In terms of written fluency (measured by the number of words produced per minute), again it was the DW participants who performed better, although the difference was minimal (P&P= 10.30; DW= 12.97, small effect size, $d = 0.567$). In line with previous research (Baraoui & Knouzi, 2018; Whithaus et al., 2008; Wolfe et al., 1996), DW seems to allow students to produce texts at a faster rate and, therefore, generally longer in terms of word count. One of the main reasons behind this increase in written fluency can be attributed to the access to a keyboard, particularly for students with a technological background, who have the skills and, more importantly, the experience in using such devices in their everyday lives[1]. Research focusing specifically on keyboarding skills and L2 writing has found evidence confirming that, although these skills may play a role in L2 written production, the relationship is substantially weaker than when compared to variables such as L2 proficiency and writing ability, for example (Barkaoui, 2013). This might explain why the differences in fluency for the DW and P&P group were minimal. In addition, with the growing incorporation of technology use in the classroom and the increase in student's writing in online environments throughout their education cycle, the issue of whether or not a student's keyboarding skills may be detrimental to their L2 written production will most likely diminish over time.

Regarding the complexity of the texts produced, minimal differences were found between the two writing conditions, both for lexical and syntactic complexity. The absence of significant differences in terms of lexical complexity are in line with Vasylets et al., (2022), but contrast with results in both Baraoui & Knouzi (2018), in which participants used a wider

[1] It is important to note that the participants in the computer-mediated writing group had recently switched to online learning due to the COVID-19 pandemic outbreak and, thus, were taking all of their classes and completing all of their coursework in an online setting.

range of vocabulary with more sophisticated items when writing on pen-and-paper, and in Chambers (2008), in which participants produced texts with more lexical variation when writing on pen-and-paper. For syntactic complexity, very minimal differences were found, with only the number of dependent clauses measure showing slightly higher differences in which the DW outperformed the P&P group, with a medium effect size (DW= 0.50; P&P= 0.44). These findings are in line with those reported in Vasylets et al., (2022), in which participants also included a greater number of dependent clauses per clause in the DW condition. These minimal differences in lexical and syntactic complexity are most likely attributed to the task in the current study. As mentioned in previous sections, the task selected was the complex version of the fire-chief task (Gilabert, 2005, 2007) which consists of a problem-solving, picture-based writing activity. Due to the nature of the task, the written output is relatively controlled, as the task includes a range of specific characters in need of rescuing and a number of situations that need to be resolved (i.e., a pregnant woman and her children, a blocked lift, only one fire engine, etc.). Therefore, in terms of lexical variation, it is likely that writers would use a similar range of lexical items, corresponding to the visual prompts in the specific task, thus leading to very similar results across writing environments for this dimension of complexity. Therefore, it is an empirical question whether other types of tasks would elicit more sophisticated and varied lexical features, and future research would benefit from testing different task types in both P&P and DW conditions in order to view potential effects of task-related variables on lexical, and syntactic complexity across writing environments. As stated in Coyle, Nicolás-Conesa, and Cerezo (forthcoming/2023), the ever-growing increase in the use DW in the language classroom "opens up a gap in the field […] complicated by the appearance of new digital genres". This increase not only calls for more research into new writing tasks and genres (blogs, wikis, etc.), but also requires studies on new feedback types, specifically tailored to emerging digital tasks (multimodal composing, for example. See Elola & Oskoz, 2022; Lee, 2022; Manchón & Coyle, 2022).

Additionally, the variable of writing environment, as defined in this current study, should be explored with larger populations to investigate the extent to which different writing environments may affect overall writing quality. More precisely, research investigating the use of digital resources whilst composing may be able to shed light on the apparent advantages of DW in improving the quality of written texts, or by providing insights into the ways in which students make use of external resources available in DW conditions. This is an incipient line of L2 writing research and the findings in the current PhD dissertation point to relevant empirically- and pedagogically-relevant questions worth addressing in future research.

VI.2. RESEARCH QUESTION 2 (RQ2). EFFECT OF PROCESSING CONDITIONS ON WRITTEN TEXTS ACROSS WRITING ENVIRONMENTS

The second research question explored potential effects of three feedback processing conditions, namely, processing WCF via(i) written languaging; (ii) think-aloud protocols; and (iii) simultaneous written languaging and think-aloud protocols, on the characteristics of the texts written (as manifested in the accuracy, fluency, and complexity of the written products).

Results showed that in terms of the accuracy, it was the simultaneous WL + TA group who had the largest reduction in errors from the pre-test to the post-test, followed by the WL only group and, finally, with the lowest global error reduction, the TA only group. These results point not only to the superiority of WL+TA feedback processing in enhancing accuracy, but also to the crucial role that WL may play in promoting text improvements, a plausible conclusion based on the observed superiority of the WL group over the TA group in the two writing environments. In line with previous research, both the think-aloud protocols (Adrada-Rafael & Filgueras-Gómez, 2019; Caras, 2019; Kim & Bowles, 2019; Qi & Lapkin, 2001; Sachs & Polio (experiment 2), 2007; Swain & Lapkin, 2002; Wigglesworth & Storch, 2012a) and the written languaging tables (Cerezo et al., 2019; Manchón et al., 2020; Moradian et al., 2017, Suzuki, 2012, 2017) used in isolation led to improvements in accuracy with no detrimental reactive effects found for either instrument. This contrasts with findings in Sachs and Polio, 2007 (experiment 1), in which the TAs were found to be negatively reactive, although in their case the feedback provided were reformulations, in contrast to the direct error correction in our study. In our research, results can be interpreted as suggesting that, the very act of processing the direct WCF provided, regardless of the instrument used to process it, provided an opportunity for the participants to reflect on the errors they had made and incorporate improvements into their post-test writing.

Looking at the writing environment variable, it is worth mentioning that for all processing treatment conditions, it was the P&P writing groups that experienced a higher overall decrease in errors from Time 1 to Time 2. In contrast to previous research (e.g., Tafazoli et al., 2014) which found that DW induced a higher number of revisions in texts written online, the data in this thesis shows that the P&P writing environment led to higher instances of feedback incorporations. However, it is important to mention that in Tafazoli et al's., (2014) study the participants were not explicitly asked to engage in feedback processing and, therefore, the discrepancies in results are more than likely related to this crucial difference between the two studies. Therefore, results in this case show that the participants who

processed their feedback on pen-and-paper benefitted more in terms of overall accuracy than those who completed WCF processing online. Research question three will provide more in-depth details on the differences found for each processing instrument for the two writing modalities investigated.

All participants, regardless of WCF processing condition, improved their overall text accuracy, as well as fluency showing once again that the WCF processing instruments used in the study were beneficial for L2 writing measures. As was the case for accuracy, the think-aloud only group was the one that experienced the lowest increase in fluency. In contrast, the participants in the WL only group were not only the ones who engaged in higher levels of processing, but also the ones whose texts showed the largest increase in fluency, both when writing on paper and on the computer. Importantly, screen-based writing resulted in higher improvements in fluency for all WCF processing conditions. As previously discussed, the availability of a keyboard in the DW condition seems to have facilitated the ease and speed at which participants produced their written texts, leading to a higher fluency overall, when compared to the texts written on paper.

In terms of complexity, results for lexical complexity -including lexical sophistication (LS2), lexical density (LD) and lexical diversity (UBER)- showed minimal differences from Time 1 to Time 2, with the feedback processing conditions not having much effect on the lexical complexity of the written texts. The only truly significant (large effect sizes) differences found corresponded to the UBER scores (lexical density) for the WL + TA groups in the DW environment. In contrast, the P&P texts showed a reduction in UBER score. These findings could be explained by the fact that the initial texts written in the DW condition had a lower number of initial errors, yet a higher percentage of lexical errors (28%) as compared to the texts written by the P&P group (18%). Therefore, the correction of these errors in their rewritten texts appears to have led to this slight increase in the UBER measure at Time 2.

For syntactic complexity, including mean length of t-units (MLT), mean length of clauses (MLC), coordinate phrases per clause (CP/C), dependent clauses per clauses (DC/C) and complex nominals per clause (CN/C), as with lexical complexity, results show minimal differences between Time 1 and Time 2 for all three WCF processing groups. However, worthy of mention is the dimension of complex nominals per clause (CN/C), which showed slightly higher reductions in the scores from Time 1 to Time 2 more than any other measure of syntactic complexity. This reduction was in the WL+TA processing condition in DW and P&P writing, as well as for the WL only processing group in DW. This might be interpreted as suggesting that writing was negatively affected by the provision and processing of feedback, albeit to a

minimal degree. Perhaps this could be attributed to the fact that the participants in the WL+TA processing condition focused more on the error corrections provided (which primarily consisted in feedback for accuracy), as manifested in their higher accuracy in Time 2, when compared to the participants in the other two processing conditions. Therefore, despite these reductions in syntactic complexity for the WL+TA condition, the overall accuracy of their written production, as measured by number of errors, was higher and proved to be the most beneficial feedback processing condition globally speaking, confirming that a combination of written languaging and think-aloud protocols may be the most beneficial for promoting language learning via feedback processing.

VI.3. RESEARCH QUESTION 3 (RQ3). EFFECT OF PROCESSING CONDITIONS ON DoP OF WCF ACROSS WRITING ENVIRONMENTS

The third research question set out to explore the relationship between the type of feedback processing engaged in by participants (i.e., written languaging, think-aloud protocols, and simultaneous written languaging and think-aloud protocols) and the levels of depth of processing that were induced by these three processing conditions, in both digital and traditional writing modalities. First, the results from the written languaging data will be discussed, followed by the think-aloud data.

VI.3.1. Written Languaging Tables

VI.3.1.1. Noticed Errors

Our data provides additional support to the results in Cerezo et al. (2019), who found that the noticing of errors was enhanced by the availability of WCF. Yet, our data shed further light by showing which WCF processing conditions led to higher error noticing, which in our case corresponded to the WL and the WL+TA groups. Our data also adds to previous research by shedding light, perhaps in a pioneering way, on the influence of writing environments on error noticing, which in our data was distinctively associated with higher levels of noticing in the DW condition. Such higher error noticing in the digital modality could be attributed to the saliency of the error corrections on the page. That is, in a digital format, the visibility of the error corrections tends to be more salient, as tools such as track changes, alongside the opportunity to change font styles, allows instructors to make a clear mark of where the errors

are in the text. In our study and, as shown in Figure 38, the feedback included (i) underlining, (ii) change in the font colour (pink), and (iii) a comment in the margin with the change made to the text. In addition, as the corrections were provided on *GoogleDocs*, any changes made to the written texts were also visible in a list format, in the version history tool that this application includes. In contrast, the errors in the P&P written texts were marked (underlined) using a different colour pen to the one used for elaborating the text (in this particular study, a red pen was used, see Figure 39), and corrections were provided above the errors, in the same colour pen. However, unlike in the DW condition, no indication of the error corrections was included in the margins of the P&P written texts, nor was there access to a list of the modifications made to the text. These differences between modalities appear to have played a beneficial role on the noticing of errors for the DW condition.

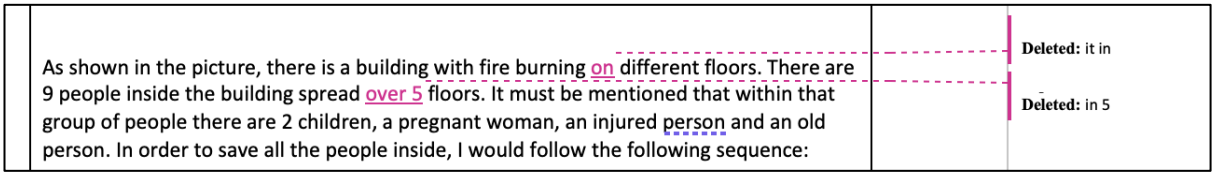


Figure 38. Example of Computer-Mediated Feedback Provision

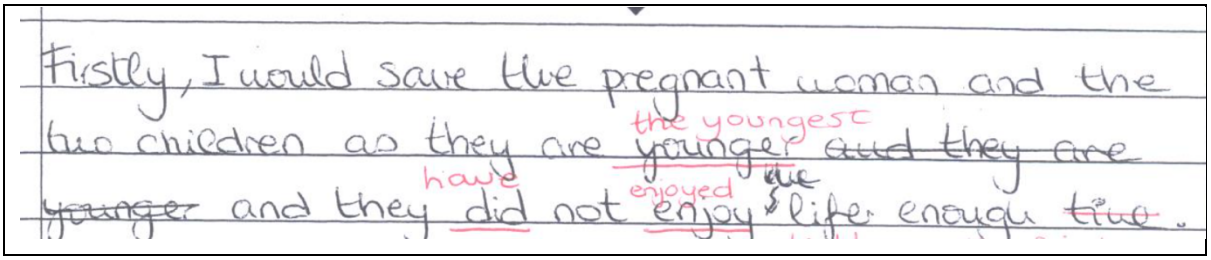


Figure 39. Example of Traditional, Pen-and-Paper Feedback Provision

As discussed in II.1.2.1., for written corrective feedback to be converted into intake, it must be noticed by learners (Schmidt, 1990). In this study, as direct WCF was provided, not only did participants have an indication of the error (underlining) but they were also provided with the correction. By providing learners with this type of salient feedback, noticing is more likely to take place. Despite a lack of research specifically looking into written corrective feedback saliency, an extensive amount of research has been carried out relating to input enhancement in other fields, particularly in the area of vocabulary acquisition. Results have confirmed the beneficial effects that increasing saliency of target items can have on L2 acquisition. For example, in Vu and Peter's (2020; 2023) studies, the treatment conditions including input-enhancement led to higher learning gains for incidental collocations in reading

modes. Further research should pursue more nuanced understandings of the way in which feedback salience (in and by itself and as mediated by composing media) may have an effect of L2 writers' noticing of their errors and of the corrections provided on them. Our data simply points to this possibility and, as such, it points to a relevant item in future research agendas.

Focusing now on the two treatment groups that included a WL table, in both writing modalities, it was the simultaneous WL+TA group who demonstrated a higher frequency of noticed errors. This finding could be attributed to the much higher implication (in terms of task time) of the participants in this treatment groups compared to the WL and TA only groups. As stated in the coding of the data, and in line with the definition of DoP provided by Leow (2015), a student spending time processing a target item, as was the case for the WL+TA groups, was interpreted a high level of DoP. Therefore, the very act of having to discuss their errors out loud whilst simultaneously writing about them in the WL table, appears to have led participants to notice a higher number of errors in both writing modalities and, in most cases, led to deeper levels of processing as a result.

Another tendency observed in the WL + TA group in the DW condition was the manner in which they approached the processing task. Many participants faced with writing and talking about the errors at the same time, opted for first noting down all of the errors in the WL table, and then going through the text again, error by error, this time orally, in order to discuss the corrections provided for each one (see example [11]). Others opted for reading the whole text first out loud, and then went back to discuss the errors individually one by one. Not only did this approach led these participants to spend more time on the processing task, as mentioned earlier, but it also meant that the participants were more aware of the error corrections in their texts and, on various occasions, they were able to notice errors that they had not noted down in their first reading of the corrected text.

[11]

Ok so the first hmm…mistake was I wrote go…instead of get out (4) then…I wrote… big…and it's more like…large (4) maybe it is more…like right now…more like…writing (22) [Writes in the written languaging table] wait…I am gonna write all the mistakes and then I can talk (3) hmm [Writes all the mistakes in the WL table (74)]

VI.3.1.2. Error Codes

The second element included in the written languaging tables was the error code section, in which participants were asked to provide a code corresponding to the type of error made (i.e., grammatical, lexical, spelling, punctuation or other). Here, the results were very similar, with both writing conditions (DW and P&P) showing comparable frequencies. However, when focusing on the processing conditions individually, it was the participants from the WL only group who correctly identified a slightly higher number of codes for the noticed errors (See Table 25 in section IV.3.1.2). This modest increase in frequency for the WL only group could be attributed to a trend found in the WL+TA data, in which participants tended to focus more on discussing the errors orally, particularly when it came to providing explanations. This focus on the think-aloud part of the WL+TA processing condition seems to relegate the written languaging table to function solely as a guide and, thus, appears to have played a detrimental role on the information provided in the table. However, and as we will discuss in the following section, the missing information was always provided in the oral think-aloud protocols.

In terms of the specific codes provided, and whether or not they were correct, the majority of incorrect codes were related to the misuse of prepositions, which students coded as lexical errors as opposed to grammatical ones (an example of which can be found in Figure 40 below). In the qualitative analysis of their data carried out by Park and Kim (2019), it was found that when learners addressed the wrong error type, they were more likely to unsuccessfully incorporate the corresponding revisions in their rewritten texts. Many of the errors that were made again in the rewritings corresponded to cases where the participants had not understood the nature of the errors and, therefore, provided the wrong error code. However, this trend was not found in our study and, despite some participants providing an incorrect error code, they were still able to successfully correct the detected error in their rewritten texts. Worthy of mention is the difference in feedback types between the two studies: The participants in Park and Kim's (2019) study were provided with indirect WCF, in contrast to the direct WCF provided in our study. The more explicit feedback type in the present study might have enabled participants to incorporate the correct revisions, despite not fully understanding the error type, as they were provided with the correct version of the target item in their feedback. In addition, the texts that were analysed after the treatment, consisted in immediate re-writings of the task, therefore, a very minimal amount of time had passed between the processing of errors and the re-writing of the texts. Thus, in the short-term, direct WCF led to a higher written

accuracy in text revisions. Yet it remains to be seen if these improvements last in a more longitudinal manner, in delayed post-tests, which is an empirical question for future research.

	ERROR	CORRECTION	CODE	EXPLANATION
1	OF IT	DELET E	L	The error could be an error in the use of the preposition.

Figure 40. Example of Wrong Error Code (taken from WL+TA DW Environment)

VI.3.1.3. Explanation of the Error

The differences found for providing a correct error code were relatively minimal when comparing both writing conditions. However, for the last column of the written languaging tables, which corresponded to providing an explanation of the error, the differences found between writing conditions and treatment groups were more noticeable. As in the case of error codes, it was the WL only group who once again provided a higher number of explanations for noticed errors in the WL table, especially in the DW condition (See Table 26, section IV.3.1.3). The lower number of explanations in the WL +TA treatment groups can be attributed to the already mentioned tendency of leaving blank spaces in the WL tables, although reflecting on the errors orally in the TA data. The example below (Figure 41) shows how a participant from the WL+TA experimental group in the P&P writing condition, includes information on the error, the correction, and the code, but does not include any information in the explanation column.

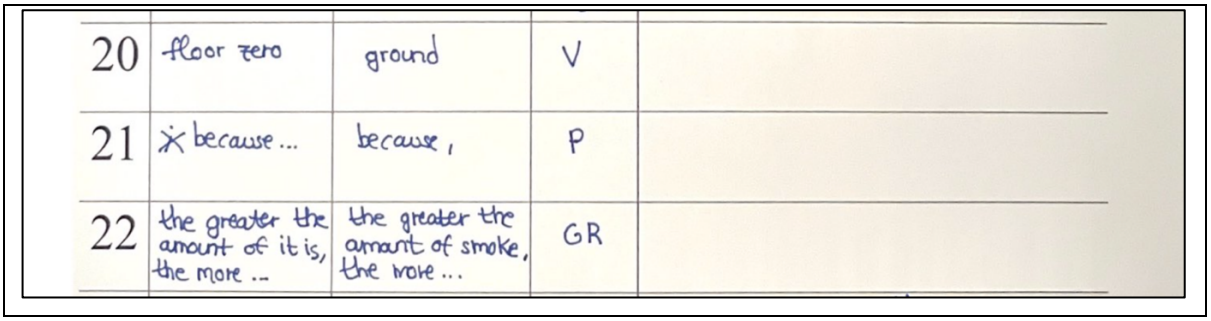


Figure 41. Example taken from WL+TA Group (P&P Writing Environment

Without consulting the corresponding TA protocol (visible below in Figure 42), which was completed simultaneous to filling in this WL table, it is impossible to know whether the participant was able to provide an explanation for the errors and, more importantly, whether this explanation was correct or incorrect. Thus, at first glance it appears to be an incomplete processing of the error, with the participant not being able to provide an explanation. However, the WL table represents only a portion of the participant's processing activity. The TA clearly shows this as the participant is able to discuss why "floor zero" has been changed to "ground floor".

TA Transcription: original	**TA Transcription: Translation**
El siguiente error es de vocabulario yo he escrito [Laughs] floor zero hmm…para referirme al primer…a la planta baja...la prim…no…no la primera planta si no (2) la planta…que esta…al entrar [Laughs] del edificio y seria…ground (15)	The next error is vocabulary I wrote [laughs] floor zero hmm...to refer to the first...to the ground floor...the fir...no...not the first floor but (2) the floor...that is...in the entrance [laughs] of the building and it would be...ground (15)

Figure 42. TA Transcription Taken from WL+TA Group (P&P Writing Environment)

As previously mentioned, this finding might be interpreted as suggesting that the participants used the WL table as a guide to complete the TA protocols, providing less information in the WL table itself, but successfully addressing this missing information in their think-alouds. In fact, there are some examples (as shown in Figure 43) in which participants did not provide any information in the table, but did successfully process the error orally, commenting on all four columns expected to be included in the table (error, correction, code and explanation) and demonstrating that, although the table was left blank, they were very much aware of the information they were required to provide and commented on all four obligatory elements.

TA Transcription: original	TA Transcription: Translation
El siguiente error es similar al que he cometido antes yo…hmm…pongo the one…who gets to the floor hmm…refiriéndome a un (3) a group…who [whispers gets to the floor] (14) yo escribí the one who gets to the floor pero es un grupo distinto del que hemos hablado anteriormente porque estamos hablando del grupo que va a salvar a…a la persona mayor que hay en el...el…en el (2) piso tercero…así que debería ser a group y no the one (8) un error similar al anterior…también de gramática (10)	The next error is similar to the one I made before...hmm...I put the one...who gets to the floor...hmm...referring to the (3) a group...who [whispers: gets to the floor] (14) I wrote the one who gets to the floor but it is a different group from the one we spoke about before so we are talking about the group that is going to save the...the elderly person that is on the...the...on the (2) third floor...so it should be the group and not the one (8) an error similar to the one previously made...it is also grammatical (10)

Figure 43. TA Taken from the WL+TA Group (P&P Writing Environment)

This tendency leads to question why the participants felt comfortable discussing the explanation of the errors orally but not in written form. Perhaps providing students with a written space in which tangible evidence of their L2 knowledge is represented can be discouraging. As suggested by Roca de Larios (personal communication, June 2022), less information could have been provided in the WL tables, particularly with respect to their doubts and uncertainty for specific L2 forms, due to the permanent trace that the written form leaves. The tangible nature of written languaging also demands higher cognitive engagement (Suzuki et al., forthcoming/2023). This is due to the fact that writers have more time to reflect on their errors as they have a physical representation of them on paper, as opposed to more immediate responses provoked by thinking aloud. As this practice is rarely implemented in L2 classrooms, especially when compared to oral languaging, Suzuki et al., (2023) suggest the relevance of providing students with the necessary training and subsequent practice before using this type of activity in classrooms/research.

VI.3.1.4. Levels of Awareness

The quantitative analysis of the levels of awareness of the participants when processing the feedback (See Table 6, based on Cerezo et al's. 2019 coding scheme explained in section III.5.2.1), showed that the participants in the DW condition performed better than the P&P group as they were the group that provided a slightly larger frequency of higher levels of awareness (particularly for level 5). This analysis is based on the amount of information provided in the written languaging tables (a representation of the outcome of the feedback processing stage) and it is important to reiterate that it did not provide a true reflection of the nature of the explanations provided. Thus, in order to explore more fully the levels of awareness of the participants, it was essential to analyse the annotations from a qualitative perspective, and hence be able to compare the information provided in the WL tables and in the think-aloud protocol data. The explanations were analysed in terms of whether or not a participant engaged with the feedback and, if so, in what ways they did. The data was therefore divided according to the categories found in Table 30 below (taken from section III.5.2.3) and complemented with qualitative examples taken from the TA protocols in order to also compare the affordances of both introspective measures.

Table 30. Types of Engagement with WCF as Manifested in the WL Tables

		As manifested in the WL
Participant does not engage with the EC		1. Leaves a blank space
		2. Annotates the error and correction with no further analysis
Participant engages with the EC	a) Disagrees with the EC	1. "I disagree" in the explanation section, but no further information provided.
		2. Provides a written explanation for their disagreement

	b) Agrees with the EC	1. Explanation section left blank or evidence of not being able to provide an explanation
		2. Evidence in the explanation section that a new L2 form has been accepted. "I didn't know that word/form"/ "I didn't know how to say it"
		3. Translation included in the explanation section
		4. Personal reasons (i.e., rushing, always make this mistake.)
		5. Rule explanation/Formulation included in the explanation section

- ***Participants not engaging with the WCF***

A lack of engagement with the EC provided was manifested in two ways: (i) a complete lack of information provided in the table (all columns left blank); or (ii) an annotation of the error and/or correction but no further information provided (no error category nor explanation). This lack of engagement in both forms corresponds to the analysis previously discussed in sections V.3.1.1 (noticing of errors) and V.3.1.3 (explanation of errors) in which the participants included blank spaces in their WL tables, and for which qualitative examples have been provided. Therefore, in order to avoid redundancy, the specific data on lack of engagement with the WCF will not be discussed here again.

- ***Participants engaging with the WCF***

The participants' engagement with the WCF provided was divided into two main categories: (i) disagreement with the EC, and (ii) agreement with the error correction.

- *Disagreement with the error corrections*

Focusing first on disagreements in the EC category, it was surprising to see a range of examples in which participants explicitly disagreed with the error correction provided. This disagreement was manifested in two manners: either with some sort of indication in the written languaging table that they disagreed with the correction provided ("I disagree") but no further explanation being given and the second, by providing an explanation as to why they did not believe the error correction was necessary.

Results indicate that participants in the P&P group were more prone to disagreeing with corrections without providing a reason why than those in the DW environment (for which no examples were found), although the frequencies were relatively low (0.45% in the WL condition and 3.22% in the WL+TA condition). Furthermore, in the WL+TA processing condition, data from both writing environments showed examples of disagreements with explanations (DW with 2.56% and P&P with 0.57% frequencies). As an example, Figure 44 shows data by a participant in the WL only group when writing in P&P. This participant simply mentions that she thinks both words fit but does not provide any further information as to why. In her writing, the participant had used "one" to refer to a person that had not been previously mentioned in the text, but she does not believe this constitutes an error. This participant went on to incorporate the corrected error in her re-writing, despite mentioning that she thought both words were appropriate.

	Error	Correction	Code	Explanation
1	One	Person	V	I think both words fit

Figure 44. Example of Disagreement from a Participant in the P&P Writing Environment (WL Only)

In Figure 45, however, taken from the DW data from the WL+TA group, the participant provides more of an explanation as to why s/he believes the correction is not necessary. In this case, because the participant wanted to express that s/he was going to do something with certainty, rather than a possibility. However, throughout the writing, this participant had used modal verbs consistently, hypothesising about the various actions they would take in order to save the characters from the burning building. In the specific sentence that was corrected, the participant used the second conditional, requiring the use of the modal verb "could" as opposed

to “will”. Despite this, the participant states that as s/he was referring to an action they were going to do, and there was no other possibility, the “will” form was more appropriate. As in the previous example, the participant still corrected the form despite not agreeing with the error correction provided.

	Error	Correction	Code	Explanation
1	Will	Could	Grammar	I was going to do it, like there wasn’t a possibility.

Figure 45. Example of Disagreement from a Participant in the DW Environment (WL+TA)

Some reasons behind these disagreements may be related to the participants’ high level of L2 proficiency (between a B2 and C1 level of English according to the CEFR). Thus, we can hypothesise that the participants were fairly confident in writing in the L2. In addition, it is important to remember that the participants were undergraduate students from a degree in English philology, specifically, in their last year (4th year) and were enrolled in an Applied Linguistics course. Another potential reason behind the confidence of the participants when processing their feedback may be related to the instructors who completed the data collection, who were two junior researchers. This may have played a role on the way in which the participants engaged with the error corrections provided as researchers and participants were close in age. To be close in age. However, this is just a speculation and not a conclusion derived from the data.

The exit questionnaires, which were administered upon completion of data collection, contain answers that may suggest why certain participants were more prone to disagreeing with the feedback than others. The participant in Figure 44 was extremely confident when writing in the L2, confirming that she had no difficulties when composing texts, she did not suffer from any type of writing anxiety, nor did she feel less confident when writing in English. She confirmed that writing in the L2 was a positive experience for her and she was more than happy to compose texts both in the L1 and L2, with equal success. These answers suggest that the participant was an extremely confident L2 writer, which was also evident in her feedback processing, in which on numerous occasions, she explicitly stated that she did not believe the

correction was necessary or she did not know why it had been corrected, as her original version appeared to be correct.

Despite a limited amount of research focusing specifically on student's beliefs towards WCF and how this may affect their uptake of target items (Rummel & Bitchener, 2015), a number of WCF studies have provided evidence on how engagement with WCF may be affected by student's beliefs (Koltovskaia & Mahapatra, 2022; Storch & Wigglesworth, 2010; Swain & Lapkin, 2002). Examples in this research show how students may reject feedback if they believe it goes against what they already know, if they believe that their intended meaning has been compromised by the WCF, or if the feedback contradicts their beliefs. These types of actions have been defined as behavioural traits of student engagement (based on Ellis's [2010] conceptual framework) and correspond to instances in which a student (i) revises or rejects the feedback received; (ii) spends time engaging with the feedback; and (iii) uses strategies to improve writing quality (Koltovskaia & Mahapatra, 2022). What previous research has found is that, depending on the student's beliefs towards WCF and the type of WCF they receive, their behaviour towards the corrections they receive will fluctuate. Therefore, the rejection of feedback examples found in the data of this doctoral thesis may not solely be based on internal factors (such as the expectation of students or their prior L2 knowledge) but also external aspects (as previously suggested), such as the instructors who provided the feedback, as well as the type of feedback provided. For example, recent research has found comprehensive feedback to be overwhelming for students (see Koltovskaia & Mahapatra, 2022), whereas contrastingly, incomprehensive feedback was found to be too confusing (see Saraghi et al., 2021). Gaining prior information on student's beliefs and perceptions towards WCF, and their feelings towards the instructors who provide such feedback, could be a way in which to avoid instances of disagreement, by specifically adapting the feedback to the students' needs and preferences.

- *Agreement with the error correction*

In cases in which participants engaged with the feedback and manifested no signs of disagreeing with the corrections provided, five categories were elaborated corresponding to the outcomes found in the data output (Table 30 aforementioned) including: (1) The explanation section left blank or with signs of not being able to provide a solid answer; (2) evidence that a new L2 form had been accepted; (3) translation; (4) personal reasons; and (5) rule explanation/formulation.

(1) The explanation section left blank or with signs of not being able to provide a satisfactory answer

The first category consisted in participants explicitly writing, in the explanation column of the WL table, that they did not understand the correction provided, or leaving a blank space in the explanation section, when all other columns had been completed correctly. The P&P WL only group showed a notably larger frequency of this type of engagement: 10.34% of noticed errors versus 4% frequency in the other treatment conditions. In Figure 46 below, the participant clearly expresses that s/he is not sure why the correction "any" is necessary but does not provide any further information or signs of processing. As this participant did not participate in the think-aloud protocols, it remains unclear whether or not s/he attempted to understand why the correction was necessary. Interestingly, the error was corrected in the rewriting perhaps showing s/he had successfully noticed the target form, despite not understanding it. As previously mentioned, the rewriting did not correspond to a delayed post-test and therefore it is difficult to assess with any degree of certainty whether or not the participant would have been able to retrieve this information at a later date.

	Error	Correction	Code	Explanation
1	Grow bigger	Grow any bigger	V	I do not really know why "any" is necessary

Figure 46. Example Taken from WL Only Group (P&P Writing Environment)

In Figure 47, we can see an example from the DW, WL+TA, and once again, the TA proves essential in being able to gain a better understanding of the explanation provided in the WL table. In the WL the participant states that s/he does not know why the error is wrong and mentions that in Spanish, the two words are not distinguished.

	Error	Correction	Code	Explanation
1	Big	Large	V	I don't know why it is wrong. In Spanish we don't

	distinguish them

Figure 47. Example Taken from WL+TA Group (DW Environment)

In the corresponding TA (Figure 48) it can be seen how the participant thinks about the distinction between the two words in English, and despite not reaching a firm conclusion on the uses of each word, the participant is able to recognise that there is a clear difference between the two and this difference might not necessarily be present in his/her L1, Spanish. The participant successfully incorporated the target item into his/her rewritten text, suggesting an understanding of the error correction provided.

TA Transcription: original

Ok…big and large…hmm…this would secure a…large exit (2) big…hmm (4) I think it's the same like in Spanish we…we…every time we want to say that something is big…we just use the verb…like the…the adjective big or (2) like…like in English we have more…more variety of…adjectives that indicate hmm…the size of something…so (2) so I think I know why…like this is also being grammatical…right?
Hmm…spelling…grammatical…punctuation…hmm…and (3) vocabulary (2) probably vocabulary (3) hmm (4) [writing in the written languaging table] …I know why it is wrong…and…maybe…and…hmm…in Spanish…we do not distinguish… them…I think…I think it is because of that…I think…in my mind (3) anyway [SIGHS] hmm…

Figure 48. Corresponding TA Transcription for the Error Processed in Figure 47

(2) Evidence that a new L2 form had been accepted

The second category of agreement corresponded to participants showing signs of a new item being learned, as manifested by statements such as: "I didn't know that word/form" or "I didn't know how to say it". The data from the DW, WL only group showed a higher frequency in this category (14.72% of noticed errors), as compared to the data from P&P writing group (only 1.66% of noticed errors). Additionally, the WL+TA groups in both writing environments showed relatively similar frequencies (4.45% in DW and 6.85% in P&P) of new item learning. The examples provided show that the participants were not aware of how to say the target item in their L2 and accepted a new L2 form. Once again, not much information is revealed

concerning the new form in the WL tables nor whether the student had understood the meaning of the new item, as shown in Figure 49 below.

	Error	Correction	Code	Explanation
1	Are open air	Have fresh air	GR	Didn't know the correct collocation

Figure 49. Example of New L2 Item Learnt in WL Only Group (DW Environment)

In contrast, and as previously mentioned as a common trend in the data, the TA protocol data reveal more information. In Figure 50 below, the student writes in the WL table that s/he didn't know how to say "walking stick" in English. In the corresponding TA (Figure 51), we can see how the student notices the target item and actually shows signs of frustration towards the error s/he had made, as manifested in his/her use of swearing and sighing. The participant mentions that s/he did not know how to say the L2 form and, accordingly, s/he made up a sentence to convey what s/he wanted to say, by describing the item ("a device he is using to facilitate walking"). The participant accepts the new term and incorporates it successfully in the post-test rewriting.

	Error	Correction	Code	Explanation
1	Device he is using to facilitate walking	Use of a walking stick	V	I didn't know how to say "walking stick" in English

Figure 50. Example from P&P Writing Environment (WL+TA)

TA Transcription: Original	**TA Transcription: Translation**
The device he is using to facilitate his walking (3) [copies down the errors into the column: device to… fa-ci-li-tate…his walking] bueno…vaya m****a desastre voy a hacer aquí… he is using…to facilitate his walking (3) [SIGHS] aquí es que no sabia como se decía…bastón…y entonces pues me invente un (2) escribí…cuando	The device he is using to facilitate his walking (3) [copies down the errors into the column: device to… fa-ci-li-tate…his walking] well...what a s**t disaster I am going to make here...he is using...to facilitate his walking (3) [SIGHS] here I didn't know how to

se que es walking stick es mucho mas…bueno este es de vocabulario…[writes the correction: use…of…walking…stick] (2) claro…a ver…[writes her explanation: I…didn't…. know…walk… how to say…how to say…walking stick…walk-ing…stick…in English] vale guay (3)	say...” bastón”...and so I just invented...(2) I wrote...when I know it is walking stick then it is much more...well...this is vocabulary…[writes the correction: use… of… walking… stick] (2) of course...let's see... [writes her explanation: I…didn't…. know…walk… how to say…how to say…walking stick…walk-ing…stick…in English] ok cool (3).

Figure 51. Corresponding TA Transcription for the Error Processed in Figure 50

(3) Translation

The next category concerned participants referring and resorting to their L1 in order to understand or explain an error. This strategy is reminiscent of Cumming's (1990) original discussion of L1 use in L2 writing to solve problems via cross-linguistic equivalents, as well as of Murphy and Roca de Larios's (2010) analysis of L1-based lexical search strategies. Our data point to the relevance of expanding previous and extensive work on L1 use in L2 writing with a fuller exploration of how L2 users resort to their whole linguistic repertoire when engaging with and making use of the feedback provided on their L2 texts. Examples of this type of L1-L2 engagement with feedback can be found in all treatment conditions. However, the data from the DW, WL only group showed a slightly higher frequency (6.76% of noticed errors), with the other groups ranging from lower frequencies of 1.00-2.50%. It is important to note that, in line with the other categories, the percentages are generally higher for the WL only group as they only had the WL tables to process their errors with. In contrast, the simultaneous WL+TA group reflected on the errors orally rather than providing written explanations in the tables provided.

In the example provided in Figure 52 below, taken from the data by a participant in the DW, WL only group, the explanation provided refers to Spanish using the same preposition for “these situations” and, therefore, the difficulty in distinguishing between the English forms “of” and “from”. There is no further elaboration on what situations the participant is referring to, nor does the participant specifically mention the context for the specific error and correction s/he is focusing on in the WL table at that time. Interestingly, this participant grouped together

a number of preposition errors and discussed them in conjunction, as manifested in the error section in which the participant states "in many occasions", referring to the recurrent incorrect use of the preposition "of" in the text. Noteworthy here is the fact that the student grouped together a range of completely different preposition errors (not just those in which "of" and "from" have been incorrectly used). In addition, and more than likely a consequence of this incorrect grouping of errors, the participant did not correct the error included in the WL table in the re-writing, and another three prepositions were unsuccessfully corrected in the revised text. Thus, despite having noticed the preposition errors, it appears that grouping them together, rather than explaining each error one by one, played a detrimental role on the languaging process of this participant, as s/he was not able to successfully incorporate the revisions into their rewritings.

	Error	Correction	Code	Explanation
1	Of (in many occasions)	From	Lexis	As in Spanish there's the same preposition for these situations, there's a difficulty in distinguishing "of" and "from"

Figure 52. Example of Comparison Between L1 and L2 from WL only Group (DW Environment)

Looking now at an example from the same writing condition (DW) but from the WL+TA group (Figure 53), the participant provides an explanation in the WL table in which s/he mentions that the error was made due to a confusion between Spanish and English rules and that s/he had seen it written like that before, so s/he thought it was acceptable. The error the participant is discussing concerns the incorrect formulation of a sentence which was missing a non-defining relative clause ("who"- "the people who are trapped"). There is no mention of the specific type of error in the WL table, just the reference to the confusion between L1 and L2 rules. Once again though, it is the TA that reveals all missing information regarding how the error was processed (Figure 54). Noteworthy is the length of the TA transcription in this particular example, which corresponds to 4 minutes of languaging just one error correction, a considerable amount of time when compared to other LRE's coded in the data.

	Error	Correction	Code	Explanation
1	*blank*	Who are	Grammatical	Confusion between Spanish and English rules, I thought it was okay to write it like that because I have seen it written like that before.

Figure 53. Example of L1 and L2 Comparison Found in the WL+TA Group (DW Environment)

TA Transcription: original

Ok so now I wrote all the mistakes…I think that I left…left one out and they are all (2) eurgh…ok so I am just gonna delete all of these (13) I am just gonna write them down…while I just correct them…hmm…ok (2) so…here…I didn't…I…I'm really…really aware of the amount…I am really aware of the amount of times I just make this mistake (4) [SIGHS] which is that I…translate from English to Spanish (3) literally…because hmm (2) in Spanish when you want to make a relative clause (2) with…with a verb to be followed by an adjective…sometimes you just take out the relative pronoun and the verb and you just…put there…you just write down…hmm…the subject or the object…and…the adjective…but you know…it's a relative clause (2) or I know in my mind…that it is a relative clause so…instead of writing help me take the rest of the people who are trapped…I just wrote hmm…help me take the rest of the people trapped (4) because I have seen it written like that…many times so…some…sometimes I am confused… whether it is ok to write it like that or not hmm…like instead of writing the relative clause you can…put the adjective first and then…the noun…so it's not a relative clause hmm…and…and…yeah but…sometimes I've…written…I've sometimes seen it written the other way round…like the noun and then the adjective…like in Spanish…the adjective order is…inversed…so…I don't know…I thought it was ok to say it like that but hmm (2) but yeah [LAUGHS] (2) ok so…hmm…grammatical (2) hmm (2) a

confusion…of…Spanish…confusion between…between Spanish and English rules [writing in the WL table] I thought…I thought (3) I thought…it…was…ok…to…write it…like that…because…I have seen it…written…like…that before (2) hmm…maybe it's just hmm…like a more poetic and it's just…more…more common in poetry and novels…and not in this context…maybe its…a…hmm (3) like a register issue (2) and…if it is like that…I would like to know so…Sophie or…Maria Dolores (2) you can…you can email me to…[LAUGHS] to tell me (2) because I am really curious about it…actually (2) hmm…yeah…so (2)

Figure 54. Transcription Corresponding to the WL Table Extract in Figure 53

The TA protocol reveals that the participant regularly makes this mistake as s/he admits to translating from Spanish to English. S/he then provides a metalinguistic explanation for the error and the correction in which s/he hypothesises about the need for relative clauses and how grammatical rules in English differ from Spanish when concerning the position of adjective and nouns ("In Spanish, the adjective order is inverted so I don't know..."). This lengthy metalinguistic discussion is only evident in the TA protocol as the WL table does not contain information concerning grammatical rules.

A common trend in the TAs from the DW group is the way in which participants directly address the researchers who collected the data, asking whether or not both options are acceptable and requesting they send an email with the corresponding explanation. No instances for this particular type of behaviour were visible in the WL tables and it was only when taking part in the TA protocols. The most probable reason for this is related to the context in which the participants took part in the study. The DW condition completed their TA protocols and WL tables in an online setting, via *Zoom* and, accordingly, they were connected to the session via webcam and microphone. Due to this, the instructor collecting the data (in this case, the author of the doctoral thesis) was present at all times throughout the data collection procedure, despite having her camera turned off for the TA protocol. This online presence clearly played a role in how the participants processed their errors and their direct remarks to the instructors "Sophie..." "You corrected..." demonstrate that they were fully aware of this. In turn, in P&P writing, this awareness in terms of the presence of instructors during the data collection procedure was not visible in the TAs nor in the WL tables and no mention of the researchers were found, despite the fact that the researchers were also in the same room as the participants when completing the data collection.

(4) Personal reasons

The penultimate category for engagement as manifested in the WL tables corresponds to participants providing personal reasons and could be related to what was just previously mentioned concerning the awareness of the presence of instructors. As found in previous research (Hoogerheide et al., 2016; Jacob et al., 2020; Lachner et al., 2018, as cited in Suzuki et al., 2023), many studies have documented the presence of personal references such as "me" and "you" when written languaging and explaining error corrections. This type of engagement was mainly found in the DW data and specifically (with a significantly higher frequency count) in the WL+TA treatment group (31.92%), in comparison to the WL only group (6.76%). This tendency to justify their errors due to personal reasons was evident in many instances in which participants substantiated their errors due to reasons such as "writing too fast", "I was in a hurry", "This was just a mistake I made in the moment", "I often make this mistake", etc. As previously mentioned, it appears that the DW condition, approached the feedback processing task in a different manner to those in the P&P writing condition. That is, rather than providing solid metalinguistic explanations and formulating rules, which is what they were instructed to do, they tended to provide personal explanations for the errors, as well as directly addressing the instructor who was collecting the data. This finding corroborates observations in González-Cruz et al., (2022) in which the authors suggested that, depending on the writing environment, "L2 writers may employ their cognitive and linguistic resources differently" (p. 626). In Figure 55 below, the participant notes down two errors which are attributed to writing too fast. The second error is also categorised by the participant as a "common error".

	Error	Correction	Code	Explanation
1	His	Him	Grammar	I wrote too fast
2	In	On	Grammar	I forgot to write it because I wrote too fast. Common error

Figure 55. Examples of Personal Reasons Found in the DW Environment (WL Only Group)

These personal reasons allow the distinction between genuine errors and "slip-of-the-pen" mistakes.[2] In general, the participants indicated a slip-of-the-pen mistake when discussing the personal reasons aforementioned. Therefore, some indicators for these mistakes included:

- "I was rushing..."
- "I wasn't paying enough attention when writing"
- "If I had re-read this, I would have known it was wrong"
- "I was writing too fast"

These indicators suggest that, assigning more time for task completion, (in this study, participants were provided with 50 minutes to compose their texts) may allow participants to revise their writing and pay more attention to these mistakes. However, given that the average time spent on task was 25,5 minutes, it appears that the mistakes are more related to the participants not taking the time to revise their texts, rather than not having sufficient time to complete the task. One solution to this issue could be to encourage students to self-correct their texts before submitting them for teacher-led feedback. This way, they would go through a filter in which any genuine mistakes (particularly those related to spelling and punctuation) would more than likely be detected by the students and therefore, corrected. In a similar way to the texts which were completed in an online setting, and thus went through a *spell check* feature, the opportunity to self-correct, in a similar way to feedback processing, would allow students to reflect on their writing, solving any language-related issues along the way. In fact, the implementation of this type of activity has been found to be beneficial in raising students' awareness and enhancing overall text improvement (Chen, 2010; Yang, 2010).

(5) Rule explanation/formulation

The final category for engagement, as manifested in the WL tables, corresponds to the participants´ attempt to formulate rules relating to the error correction provided. As opposed to the previous category, this type of engagement is much more metalinguistic in nature as participants attempt (at times successfully, at times not) to provide an explanation for the error corrections via the formulation of grammatical rules. This type of engagement is directly related to higher levels of awareness as it implies that the student has not only noticed the

[2]I would like to thank Lourdes Ortega for this personal communication regarding the distinction between genuine errors and "slip-of-the-pen" mistakes (TESOL Doctoral Summer School, Malta, 2019)

correction, but has also hypothesised about the L2 input and, in most cases, understood it correctly (Leow & Driver, 2021). Some examples of this high level of engagement are found in the data by the WL only groups, in which a high frequency of this type of engagement was observed: 47.25% of noticed errors were related to the formulation of rules in DW, 53.34% in P&P. This category included instances in which students provided examples of grammar rules, verb formations, discussed precise lexical definitions, and explained rules regarding word order. In the two examples provided below in Figure 56, a participant from the P&P condition (WL only group) stated the correct lexical meaning of "fog" which had been incorrectly used to refer to the smoke of the fire in this case. In example 2, the participant shows understanding of the error correction "least" as they were able to explain that the superlative form was more adequate in that context rather than the comparative form "less".

	Error	Correction	Code	Explanation
1	Fog	Smoke	V	"Fog" is a meteorological phenomenon, "smoke" is the correct word
2	Less	Least	GR	"Less" is the comparative form, but in this context, it should be "least", the superlative one

Figure 56. Examples of "Rule-Formation" Taken From a Participant in the P&P Writing Environment (WL Only)

As previously discussed, the WL table provides the outcome of the processing stage (Manchón et al., 2020), as clearly visible in the examples above (Figure 56). The explanation column contains the outcome of the participants cognitive processing of the feedback, as we are provided with the "conclusion" to which the participant has arrived at. However, no information on how they arrived at this rule is included and any information on the participants hypothesising rules and/or discussing other options became lost in the tables, confirming Manchón et al's (2020) conclusions that the written languaging tables can only capture part of the DoP of learners.

In order to fully capture this process, it was essential to recur to the think-aloud protocols to gain a deeper understanding of how exactly the student arrived at the conclusions they had included in the written languaging table, and, more importantly, to be able to discern levels of depth of processing. Therefore, the following section discusses the results from the think-aloud data and provides an in-depth analysis of the levels of depth of processing and their relation to the information provided in the written languaging tables.

VI.3.2. Think-Aloud Protocols

In contrast to the data provided by the written languaging tables (which provided the outcomes of the processing stage), the think-aloud protocol allowed a more direct access to the participants' cognitive processing while engaging with the WCF provided on their writing. As reported in the Results section (V.3.2.), the data was analysed according to four levels of DoP including: (i) High, (ii) Medium, (iii), Low, and (iv) Null.

Results show that high levels of DoP were related to the WL+TA processing condition, both in DW (44.44% of noticed errors) and P&P writing modalities (41.91% of noticed errors). These results contrast with those of the TA only group, who had much lower frequencies of high DoP levels (25.30% for DW and 11.72% for the P&P writing modality). As mentioned in the previous section, it appears that the written languaging table proved essential not only in guiding the students through the processing task, but also to process their errors at a deeper level. Thus, in the case of the WL+TA groups, regardless of the writing modality, the participants spent a greater amount of time processing their errors, which therefore led to the higher frequency of deeper levels of processing found in the results. In contrast, the lower frequencies for the TA only group can be attributed to a much lower amount of time on task. The difference between the average time spent on the processing task between the TA only group and the WL+TA was considerable, with the TA only groups spending an average of 4.96 (SD=1.67) and 5.12 minutes (SD=2.98), in the DW and P&P writing conditions, respectively. In the WL+TA condition, in contrast, the participants' time-on-task increased to 22.11 (SD=12.83) and 31.32 minutes (SD=3.57) in the DW and P&P writing conditions, respectively. The results found for the way in which students grouped together the errors are in line with the lower average of time spent on processing the errors. As reported in Caras (2019), the participants were found to generalise the errors made when processing via TAs and grouped them together in categories, rather than discussing them one by one. This tendency was found in the direct WCF group in Cara's (2019) research and is corroborated by the results found in

the present study. As participants were provided with the error corrections, they were able to skim over the texts, without stopping at each error individually and hypothesising about the correction (as perhaps would be the case if indirect WCF were provided, for example, and as was the case when they had a WL table present). By doing this, the TA only group demonstrated higher frequencies for low levels of DoP (57.42% in the P&P writing condition and 48.61% in the DW condition), as participants tended to simply repeat or read error corrections, agreeing with them, but not providing any further information. If any further discussion was included, it generally related to groups of errors, rather than individual ones. In line with results found in Koltovskaia and Mahapatra (2020), the grouping of errors was related to students noticing a repetition in the same errors made over and over again, for example on the use of prepositions. Generally, when this happened, students would accept the revisions as they were "quite obvious" (Koltovskaia & Mahapatra, 2022, p. 306) and make no further comments.

This tendency also explains the results for null DoP levels, in which the data by TA only group showed higher frequency, especially for the DW environment, with a total of 17.41% of noticed errors, compared to 8.79% in the P&P writing environment. These higher frequencies of null DoP compared to the participants in the WL+TA treatment group (0% for DW condition and 6.35% for the P&P writing condition), are most likely related to this tendency of participants grouping together the errors and generalising the comments made. By doing so, they avoided languaging the errors one by one and, therefore, had a higher frequency of instances lacking any information, which were classed as null DoP.

For the medium levels of DoP, the results found showed that, as with high DoP, it was the WL+TA group who showed higher instances of medium DoP in the DW environment, with 31.65% of noticed errors classified as such, in comparison to only 8.68% for the TA only group. In contrast, both processing conditions in the P&P group showed similar scores for medium levels of DoP, with 22.07% of noticed errors in the TA group and 20.55% in the WL+TA group. One reason behind the higher number of medium DoP levels in the DW condition could be related to the observed tendency by participants to provide personal, rather than linguistic, reasons for their errors. Generally, as mentioned in the previous section, participants would explain their errors by attributing them to reasons such as writing too fast or not paying much attention. By doing so, they would discuss the error and the correction, demonstrating an understanding of both but rather than provide a metalinguistic explanation (classified as high DoP), they would simply discuss why they had made the error, from a more personal point of view (see Figures 57 and 58 below).

TA Transcription: original	TA Transcription: translation
El tercer error es ask que no he puesto el (3) pronombre…hmm (3) entonces seria gramática también…y de nuevo creo que es un error por ir…escribiendo sin fijarme [writes down the error, correction, code and explanation (38)]	The third error is ask where I haven't put the (3) pronoun.... hmm (3) so it would bne grammatical too... and again I think it is an error for going...writing without paying attention [writes down the error, correction, code and explanation (38)]

Figure 57. Example of "Personal Reasons" Taken from the WL+TA Group (P&P Writing Environment)

TA Transcription: original
Ok…I always mix…hmm…in and on…so when it says close to the exit…on the first and ground floor…I always mix in and on…so that is why I made a mistake there (29)

Figure 58. Example of "Personal Reasons" Taken from TA Only Group (DW Environment)

All participants received the same instructions for the oral languaging task, which specifically asked them to complete metacognitive TAs. This decision was made so as to be able to compare the oral languaging output with the written languaging output, which was also metacognitive in nature. In theory, this pushed output (Swain, 1985) provides students with a platform in which they are required to analyse aspects of grammar in their L2, testing out hypotheses and reflecting on their written output. Unexpectedly, however, the writing modality played a role on the students' interpretations of the instructions for the metacognitive TAs and those in an online environment, working on the computer, provided much less metalinguistic information, leading to a higher number of instances of medium levels of DoP, rather than high levels.

CHAPTER VII. CONCLUSIONS, LIMITATIONS, AND IMPLICATIONS

This chapter presents the conclusions and implications of the doctoral thesis by providing global reflections on the results obtained, limitations, and suggestions for future research agendas on written corrective feedback processing in SLA-oriented L2 writing research. The last section offers a number of potential implications for research and practice.

VII.1. CONCLUSIONS

The insight obtained in this doctoral thesis has provided empirical evidence on the effects of two types of introspective measures (written languaging tables and think-aloud protocols) on both the levels of depth of processing of WCF and the effects of such processing on subsequent L2 writing accuracy. The study has additionally contributed to SLA-oriented L2 writing studies in the field of L2 writing by shedding additional light on the differences between writing in a traditional pen-and-paper environment versus a computer-mediated setting.

In terms of the implications of writing environments addressed in Research Question 1, the texts written in a computer-mediated setting were found to be more accurate in terms of overall accuracy, a finding attributed to the availability of the *spell check* function of the application used for writing the texts (*GoogleDocs*), which corrected minor errors related to spelling and punctuation. In addition, the speed at which participants elaborated their texts was slightly higher than those written in the pen-and-paper environment, as manifested in higher scores for fluency measures. This finding was attributed to the participants' skill and experience in using the keyboard given the ever-present use of computer and other electronic devices in their everyday lives. This might contrast with the practical absence of handwriting in their academic tasks and personal lives. Given the ever-growing incorporation of computers in the language classroom and the importance technology has gained over the past few decades, the positive results for writing in a computer-mediated environment are reassuring for instructors who may be faced with switching from more traditional writing settings to more modern approaches. At a research level, this finding calls for necessary caution when generalizing findings from past paper-based L2 writing research to all writing environments. Screen-based writing is likely to show its own idiosyncrasy worth exploring in future research. This research should not only zoom into digital writing processes and products but, importantly, also on comparison across writing environments given the very scarce research

that has undertaken such comparison (Cruz et al, 2022; Vasylets et al., for recent initiatives in this relevant research agenda)

In terms of feedback processing, for the population under study (upper-intermediate undergraduate students), we found that a combination of written languaging and think-aloud protocols was the optimum condition for enhancing deeper levels of depth of processing and, therefore, subsequent increased L2 written accuracy. Simultaneous WL and TA led participants to spending more time processing the WCF they had been provided with, leading to higher levels of DoP. The participants devoted more time (as compared to the time-on-task by the groups who completed the processing instruments in isolation), going through their error corrections, hypothesising, discussing rules, translating the target item and even, at times, providing a well-justified reason not to agree with the corrections. These actions, interpreted as high DoP, subsequently led to a higher number of revisions in the re-written texts.

Although it was the simultaneous group who showed the greatest improvements and higher levels of DoP, the introspective measures used in isolation also proved to be beneficial for participants. In fact, the WL only group marginally outperformed the WL+TA group when focusing solely on the information provided in the WL tables, suggesting that the completion of the WL tables in isolation may lead to high levels of awareness and engagement during the feedback processing stage. Thus, the WL tables provided a platform for which the participants could provide information on the errors they made, and according to the results obtained, guided the participants through their errors sufficiently enough to incorporate a high number of revisions in their revised texts.

The TA protocols, in turn, when completed in isolation, were also beneficial for the participants in terms of their effect on L2 accuracy. As with the other WCF processing conditions, the TAs led participants to incorporate a notable number of revisions into their revised texts, albeit to a lesser extent than the participants in the WL+TA groups. In addition to these benefits in terms of improved L2 accuracy, the TAs were also instrumental from a methodological perspective as they were able to provide rich data from which to infer the participants' depth of processing. This is taken as confirmation of previous positions on the relevance and affordances of TA protocols for DoP research (e.g., Caras, 2019; Leow, 2020; Leow & Manchón, 2021, 2022; Leow & Suh, 2022; Park & Kim, 2019).

Regarding the effects of writing environments on the processing of WCF, results showed that the processing that was done in P&P writing conditions led to greater improvements in subsequent rewritings when compared to DW conditions. In contrast, higher levels of depth of processing were observed in the DW environment. As discussed, although

the DW writing modality did not engage students in as much metalinguistic languaging as was the case in the P&P modality, processing WCF in a digital environment still succeeded in engaging participants in deep levels of feedback processing as operationalized in our coding scheme. However, the P&P writing condition appeared to be more likely to lead participants to engage in metalinguistic languaging and, as a result, obtain higher levels of accuracy in subsequent rewritings. Given the exploratory nature of our research, these findings should be taken with caution and be interpreted as tendencies to be explored in future studies. From a research methodology perspective, the potential contribution of the thesis for future work should be emphasized: the thesis is a pioneering attempt to test the affordances of diverse methodological procedures for the analysis of DoP of WCF. Future work should ascertain whether the results obtained apply to other populations of L2 writers, especially writers of lower proficiency L2 levels and, crucially, L2 writers that differ in their background in language and linguistics. Future research should also ascertain whether the results of our work and methodological implications apply to processing other types of feedback provided on the texts resulting from engagement in a range of tasks. After all, we investigated one proficiency level, one task, and one feedback strategy, as more fully discussed in the next section.

VII.2. LIMITATIONS

Despite the potential contribution of our study in the terms specified in the previous section, it is not without limitations

Firstly, the number of participants the study was lower than we would have hoped for. However, due to the COVID-19 pandemic, the collection of data became an extremely difficult task and the number of participants who were able to take part in the study was reduced to 36. Due to this low number of participants, it was not possible to include a control group and, therefore, the data collected on feedback and feedback processing effects was not compared to a group who received no treatment and yet wrote and rewrote their texts.

Another limitation relates to the participants' L2 proficiency level. As previously mentioned, the difficulty the pandemic caused for recruiting participants in part contributed to not being able to include L2 proficiency as a covariate in the study. The participants taking part in this study were all upper-intermediate or advanced learners from a language and linguistics background and, therefore, the results may only be extrapolated to this context. As mentioned in the previous section, future research would benefit from exploring WCF processing via WL

and TAs with lower-level L2 proficiency learners, from a range of backgrounds, in order to shed a stronger light on the effects of the processing conditions in focus in our research

Additionally, as pointed in the last paragraph of the previous section, the research presented in this doctoral thesis was all focused around one writing task (complex version of the fire-chief task), and one type of WCF (direct). Therefore, future research ought to explore the affordances of the introspective measures included in the study, in relation to different types of feedback and a range of different task types.

Another limitation that cannot be dismissed is related to the type of study that was completed. As described, in the Method section, the data collection was completed over the course of one-week period and consisted in a one-shot study in which the results were determined in a pre-post test design. Due to the limitations previously mentioned regarding the pandemic and the difficulties this situation caused for data collection, the decision was made to not include a delayed post-test. Therefore, the results obtained should be interpreted with caution as no evidence has been provided on the effects of WCF processing on long-term writing gains. In order to truly explore the effects of the treatment provided, and to eliminate other variables that may have played a role on the improvement in the text revisions, such as participants memorising the error corrections (Sachs & Polio, 2007), a delayed post-test would be advisable as this would contribute to a better understanding of the extent to which the writing and processing procedures implemented played a role on L2 accuracy improvements, and whether or not such potential language learning gains are mediated by composing modalities.

In addition to the relatively short data collection procedure, the study was also carried out on a voluntary basis, thus, the students who volunteered to take part in the study, were invited to do so out of class hours. In doing so, the research did not form part of the curriculum, nor did it interfere with the classes they were enrolled on. It would be interesting to explore the effects of feedback processing in a curriculum-based study, particularly relevant due to calls to apply more curricular-based perspectives in WCF research (see Leow, 2020; Leow & Manchón, 2021; Manchón & Leow, 2020). Such curricular approach would allow future research to explore longitudinally the ways in which processing feedback can enhance students L2 writing skills and language learning within the confines of what is possible in real classroom settings.

VII.3. IMPLICATIONS

The current study has implications not only for future research agendas from both an empirical and methodological point of view as discussed above, but it is also relevant for pedagogy in the field of SLA.

First, with regards to writing environments, the results provided new empirical evidence on the effects of pen-and-paper versus computer-mediated writing on L2 written production. Specifically, the study showed that the incorporation of computer-mediated writing tasks had positive effects on text quality as the quality of L2 written texts completed online, via *GoogleDocs*, proved to be more accurate than those written on pen-and-paper. It follows that by asking students to complete tasks in an online writing environment, instructors may save time on feedback provision, as texts go through a preliminary filter in which most punctuation and spelling errors are automatically corrected by the online writing tool. This beneficial feature of many DW tools allows teachers to dedicate more time to focus on correcting grammatical and lexical errors, as mechanical errors are dealt with via the tools included in the writing apps.

From a methodological perspective, the study has shed light on the data output provided by two introspective measures (written languaging tables and think-aloud protocols). Results showed that in order to truly tap into L2 writers' cognitive processes while processing feedback, oral introspective measures are more likely to provide better insights, as the think-aloud protocols were the only instrument that provided data from which it was possible to infer the participants' levels of DoP. Written languaging tables, however, despite proving useful for eliciting information on the participant´s metalinguistic awareness, failed to provide data from which to infer levels of DoP. This research implication may serve as a basis from which future studies can expand on the insights provided by the two introspective measures included, by triangulating this data with other instruments such as keystroke logging and eye-tracking devices, in order to gain further insights on writing processes and WCF processing.

In addition, the study showed that the writing environment in which the WCF processing activity is completed may play an important role on how students approach the WCF processing task. DW conditions did not play a negative role on the processing of errors from a learning perspective, but the processing the students did was much less metalinguistic in nature when compared to the P&P condition. From a methodological point of view, it is important to consider this if the ultimate aim of the research is to view the introspective

measures from a metalinguistic perspective. Our data may be interpreted as suggesting that P&P writing environments may be more effective in eliciting this type of data.

The study has contributed to current understandings of the cognitive processing of undergraduate students when provided with direct WCF. By doing so, the study has confirmed the benefits processing WCF can have on subsequent L2 accuracy, particularly when this is done via tasks such as completing a written languaging table or think-aloud protocols. In addition, if these two tasks are done simultaneously, the benefits appear to increase. Therefore, from a pedagogical perspective, it is important to bear in mind that the way in which we ask our students to process feedback will have a direct impact on (i) how they approach the processing of errors; (ii) how well they will perform on the processing task itself in terms of DoP and; (iii) on how likely their processing of errors will lead to subsequent language learning and improvements in written production. However, despite the benefits the introspective measures have on second language accuracy, we are fully cognisant that the implementation of this type of feedback processing in a classroom is not an easy task (particularly the combination of both instruments) and would require a great effort from both teachers and students alike, or slight adaptations in order to facilitate the implementations of these types of tasks within a classroom setting. Some ways in which this problem can be approached in manageable ways might include, for example, asking students to participate in group discussions, creating a teacher-led discussion in which common errors are commented on, or asking students to work in pairs and engage in collaborative dialogues. Taking into consideration the findings of this thesis, all of these options for processing feedback orally would be likely enhanced if accompanied by a written prompt, whether that involves a written languaging table or a note-taking activity, for example.

The present PhD was carried out with empirical and methodological aims in mind. The data collected has shed light not only on the role the variable of writing environment (namely, pen-and-paper versus computer-mediated) may play on L2 writing, but it has also contributed to current methodological debates on WCF processing by exploring the affordances of two introspective measures (written languaging and think-aloud protocols) and their combination. By providing insights into the cognitive processes of undergraduate students when processing feedback and showing the benefits that can result from this type of task, it is expected that this doctoral thesis has contributed further evidence to the field of L2 writing and WCF. Our research clearly shows that providing WCF for L2 learners and promoting opportunities for feedback processing provides opportunities for students to reflect on and potentially expand their existing L2 knowledge in the context of instructed SLA.

REFERENCES

Abbuhl, R. (2021). Interactionist Approach to Corrective Feedback in Second Language Acquisition. In H. Nassaji & E. Kartchava (Eds.), *The Cambridge handbook of corrective feedback in second language learning and teaching* (pp. 44-64). Cambridge University Press. https://doi.org/10.1017/9781108589789.003

Adrada-Rafael, S., & Filgueras-Gómez, M. (2019). Reactivity, language of think-aloud protocol, and depth of processing in the processing of reformulated feedback. In R.P. Leow (Ed.), *The Routledge handbook of second language research in classroom learning* (pp. 199-211). Routledge.

Ai, H., & Lu, X. (2010). *A web-based system for automatic measurement of lexical complexity.* Paper presented at the 27th Annual Symposium of the Computer-Assisted Language Consortium (CALICO-10). Amherst, MA. June 8-12.

Allen, D. (1992). *Oxford Placement Test* 2 (New edition). Oxford University Press.

Amiryousefi, M. (2016). The differential effects of two types of task repetition on the complexity, accuracy, and fluency in computer-mediated L2 written production: a focus on computer anxiety. *Computer Assisted Language Learning*, *29*(5), 1052–1068. https://doi.org/10.1080/09588221.2016.1170040

Anderson, J. R. (1982). Acquisition of cognitive skill. *Psychological Review*, *89*(4), 369-406. https://doi.org/10.1037//0033-295X.89.4.369

Applebee, A. N. (1984). Writing and reasoning. *Review of Educational Research*, *54*(4), 577-596. https://doi.org/10.3102%2F00346543054004577

Ashwell, T. (2000). Patterns of teacher response to student writing in a multiple-draft composition classroom: Is content feedback followed by form feedback the best method? *Journal of Second Language Writing, 9*(3), 227-257. https://doi.org/10.1016/S1060-3743(00)00027-8

Askvik, E., Van der Weel, F. R., & Van der Meer, A. L. (2020). The importance of cursive handwriting over typewriting for learning in the classroom: A high-density EEG study of 12-year-old children and young adults. *Frontiers in Psychology*, *11*, 1810. https://doi.org/10.3389/fpsyg.2020.01810.

Aubrey, S. (2012). *Students' reactions to using technology in an EAP writing class*. Paper presented at the 9th International Far Eastern English Language Teachers Association (FEELTA) conference, Vladivostok, Russia.

Aubrey, S. (2014). Students' attitudes towards the use of an online editing program in an EAP course. Annual Research, 17, 45-44.

Barkaoui, K. (2013). Examining the impact of L2 proficiency and keyboarding skills on scores on TOEFL-iBT writing tasks. *Language Testing*, *31*(2), 241-259. https://doi.org/10.1177/0265532213509810

Barkaoui, K., & Knouzi, I. (2018). The effects of writing mode and computer ability on L2 test-takers' essay characteristics and scores. *Assessing Writing*, *36*, 19-31. https://doi.org/10.1016/j.asw.2018.02.005

Bazerman, C., Little, J., Bethel, L., Chavkin, T., Fouquette, D., & Garufis, J. (2005). *Reference guide to writing across the curriculum*. Parlor Press.

Bitchener, J. (2008) Evidence in support of written corrective feedback. *Journal of Second Language Writing*, *17*,102–18. https://doi.org/10.1016/j.jslw.2007.11.004

Bitchener, J. (2016). The content feedback practices of Applied Linguistics doctoral supervisors in NZ and Australian universities. *Australian Review of Applied Linguistics*, *39*(2), 105–21.

Bitchener, J. (2019). The intersection between SLA and feedback research. In K. Hyland & F. Hyland (Eds.), *Feedback in second language writing. Contexts and issues* (pp. 85-105). Cambridge University Press.

Bitchener, J. (2021). Written corrective feedback. In H. Nassaji & E. Kartchava (Eds.), *The Cambridge handbook of corrective feedback in second language learning and teaching* (pp. 207-225). Cambridge University Press. https://doi.org/10.1017/9781108589789.011

Bitchener, J., & Ferris, D. R. (2012). *Written corrective feedback in second language acquisition and writing*. Routledge. https://doi.org/10.4324/9780203832400

Bitchener, J., Knoch, U. (2008). The value of written corrective feedback for migrant and international students. *Language Teaching Research Journal 12* (2), 409–431. https://doi.org/10.1177%2F1362168808089924

Bitchener, J. and Knoch, U. (2009) The relative effectiveness of different types of direct written corrective feedback. *System 37*(322–9). https://doi.org/10.1016/j.system.2008.12.006

Bitchener, J., & Knoch, U. (2010). The contribution of written corrective feedback to language development: A ten month investigation. *Applied linguistics*, *31*(2), 193-214. https://doi.org/10.1093/applin/amp016

Bitchener, J., & Storch, N. (2016). *Written corrective feedback for L2 development.*

Multilingual Matters. https://doi.org/10.21832/9781783095056

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*(3), 191-205. https://doi.org/10.1016/j.jslw.2005.08.001

Bowles, M. A. (2010). *The think-aloud controversy in second language research*. Routledge. https://doi.org/10.4324/9780203856338

Bowles , M. A., & Gastañaga , K. (2022). Heritage, second and third language learner processing of written corrective feedback: Evidence from think-alouds. *Studies in Second Language Learning and Teaching*, *12*(4), 675–696. https://doi.org/10.14746/ssllt.2022.12.4.7

Bowles, M. A., & Leow, R. P. (2005). Reactivity and type of verbal report in SLA research methodology: Expanding the scope of investigation. *Studies in Second Language Acquisition*, *27*(3), 415-440. https://doi.org/10.1017/S0272263105050187

Britton, J., & Pradl, G. M. (1982). *Prospect and Retrospect: Selected Essays*. Boynton/Cook Publishers

Bruner, J. (1983). The acquisition of pragmatic commitments. In R. Golinkoff (Ed.), *The transition from prelinguistic to linguistic communication* (pp. 27-42). Erlbaum.

Bygate, M. (2006). Areas of research that influence L2 speaking instruction. In E. Usó-Juan & A. Martínez-Flor (Eds.), *Current trends in the development and teaching of the four language skills* (pp.159-186). Mouton De Gruyter. http://dx.doi.org/10.1515/9783110197778.3.159

Byrnes, H., & Manchón, R. M. (Eds.) (2014). *Task-Based Language Learning–Insights from and for L2 Writing*. John Benjamins.

Caras, A. (2019). Written corrective feedback in compositions and the role of depth of processing. In R.P. Leow (Ed.), *The Routledge handbook of second language research in classroom learning* (pp. 186-198). Routledge.

Cerezo, L. (2021). Corrective feedback in computer-mediated versus face-to-face environments. In H. Nassaji & E. Kartchava (Eds.), *The Cambridge handbook of corrective feedback in second language learning and teaching* (pp. 494-519). Cambridge University Press. https://doi.org/10.1017/9781108589789.024

Cerezo, L., Manchón, R. M., & Nicolás-Conesa, F. (2019). What do learners notice while processing written corrective feedback? A look at depth of processing via written languaging 1. In R. Leow (Ed.), *The Routledge handbook of second language research in classroom learning* (pp. 171-185). Routledge.

Chambers, L. (2008). Computer-based and paper-based writing assessment: A comparative text analysis. *Research Notes, 34*, 9–15.

Chandler, J. (2003). The efficacy of various kinds of error feedback for improvement in the accuracy and fluency of L2 student writing. *Journal of Second Language Writing*, *12*(3), 267-296. https://doi.org/10.1016/S1060-3743(03)00038-9

Chang, C. W., Pearman, C., & Farha, N. (2012). Second Language Acquisition: Implications of Web 2.0 and Beyond. *Critical Questions in Education*, *3*(2), 52-64.

Chen, C. H. (2010). The implementation and evaluation of a mobile self- and peer-assessment system. *Computers & Education, 55*(1), 229–236.

Cheung, Y. L. (2012). Critical review of recent studies investigating effects of word processing-assisted writing and pen-and-paper writing on the quality of writing and higher-level revisions. *Procedia-Social and Behavioral Sciences*, *46*, 1047-1050. https://doi.org/10.1016/j.sbspro.2012.05.246

Chong, S. W. (2019). College students' perception of e-feedback: a grounded theory perspective. *Assessment & Evaluation in Higher Education*, *44*(7), 1090-1105. https://doi.org/10.1080/02602938.2019.1572067

Cohen, A. D. (2013). Verbal Report. In C. Chapelle (Ed.), *Encyclopedia of Applied Linguistics* (pp. 1-5). Wiley-Blackwell. https://doi.org/10.1002/9781405198431.wbeal1261

Coyle, Nicolás-Conesa & Cerezo (forthcoming/2023). Overview of methodological procedures in research on written corrective feedback processing. In R.M. Manchón & J. Roca de Larios (Eds), *Research methods in the study of L2 writing processes.* John Benjamins.

Coyle, Y., & Roca de Larios, J. R. (2020). Exploring young learners' engagement with models as a written corrective technique in EFL and CLIL settings. *System*, *95*, 102374. https://doi.org/10.1016/j.system.2020.102374

Cumming, A. (1990). Metalinguistic and ideational thinking in second language composing. *Written Communication*, *7*, 482–511.

Cumming, A. (2020). L2 writing and L2 learning. Transfer, self-regulation, and identities. In R. M. Manchón (Ed.), *Writing and language learning. Advancing research agendas* (pp. 29-48). John Benjamins

Cunningham, K. J. (2019). Student perceptions and use of technology-mediated text and screencast feedback in ESL writing. *Computers and Composition*, *52*, 222-241. https://doi.org/10.1016/j.compcom.2019.02.003

DeKeyser, R. M. (2015). Skill acquisition theory. In B. VanPatten & J. Williams (Eds.), *Theories in second language acquisition* (pp. 94-112). Routledge.

Ducate, L., & Arnold, N. (2012). Computer-mediated feedback: Effectiveness and student perceptions of screen-casting software versus the comment function. In G. Kessler, A, Oskoz, & I. Elola, (Eds.). Technology Across Writing Contexts and Tasks (pp. 31-56). San Marcos, TX: Computer Assisted Language Instruction Consortium.

Ellis, R. (2008). *The study of second language acquisition* (2nd ed.). Oxford: Oxford University Press.

Ellis, R. (2010). Epilogue: A framework for investigating oral and written corrective feedback. *Studies in Second Language Acquisition, 32*, 335–349. https://doi.org/10.1017/S0272263109990544

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. https://doi.org/10.1016/j.system.2008.02.001

Ellis, R., & Yuan, F. (2004). The effects of planning on fluency, complexity, and accuracy in second language narrative writing. *Studies in Second Language Acquisition*, *26*(1), 59–84. https://doi.org/10.1017/s0272263104026130

Elola, I., & Oskoz, A. (2016). Supporting second language writing using multimodal feedback. *Foreign Language Annals*, *49*(1), 58-74. https://doi.org/10.1111/flan.12183

Elola, I., & Oskoz, A. (2017). Writing with 21st century social tools in the L2 classroom: New literacies, genres, and writing practices. *Journal of Second Language Writing*, *36*, 52–60. https://doi.org/10.1016/j.jslw.2017.04.002

Elola, I., & Oskoz, A. (2022). Reexamining feedback on L2 digital writing. *Studies in Second Language Learning and Teaching*, *12*(4), 575-595. https://doi.org/10.14746/ssllt.2022.12.4.3

Emig, J. (1977). Writing as a mode of learning. *College Composition and Communication*, *28*(2), 122–128. https://doi.org/10.2307/356095

Ene, E., & Upton, T. A. (2014). Learner uptake of teacher electronic feedback in ESL composition. *System, 46*, 80-95. https://doi.org/10.1016/j.system.2014.07.011

Ericsson, K. A., & Simon, H. A. (1993). *Protocol analysis: Verbal reports as data*. The MIT Press.

Farahani, A. A. K., & Meraji, S. R. (2011). Cognitive task complexity and L2 narrative writing performance. *Journal of Language Teaching and Research*, *2*(2), 445–456. https://doi.org/10.4304/jltr.2.2.445-456

Ferris, D. R. (1995). Student reactions to teacher response in multiple draft composition classrooms. *TESOL Quarterly, 29*, 33-53.

Ferris, D. (1999). The case for grammar correction in L2 writing classes: A response to Truscott (1996). *Journal of Second Language Writing*, 8, 1-10. https://doi.org/10.1016/S1060-3743(99)80110-6

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 University Press.

Ferris, D. R. & Roberts, B. J. (2001). Error feedback in L2 writing classes: How explicit does it need to be? *Journal of Second Language Writing*, 10, 161–184. https://doi.org/10.1016/S1060-3743(01)00039-X

Ferris, D., Liu, H., Sinha, A., & Senna, M. (2013). Written corrective feedback for individual L2 writers. *Journal of Second Language Writing, 22*(3), 307-329. https://doi.org/10.1016/j.jslw.2012.09.009

Fox, M. C., Ericsson, K. A., & Best, R. (2011). Do procedures for verbal reporting of thinking have to be reactive? A meta-analysis and recommendations for best reporting methods. *Psychological Bulletin, 137*(2), 316–344. https://psycnet.apa.org/doi/10.1037/a0021663

Frantzen, D. (1995). The effects of grammar supplementation on written accuracy in an intermediate Spanish content course. *Modern Language Journal*, *79*(3), 329-344. https://doi.org/10.1111/j.1540-4781.1995.tb01108.x

Gao, J., & Ma, S. (2020). Instructor feedback on free writing and auto- mated corrective feedback in drills: Intensity and efficacy. *Language Teaching Research*, *8*, 1–24. https://doi.org/10.1177/ 1362168820915337

Garcés-Manzanera, A. (2022). Connecting pauses and revision behavior in children's L2 writing. In D. Moya López and A. I. Nogales-Bocio (Eds.), *Del periodismo a los prosumidores. Mensajes, lenguajes y sociedad* (pp. 289-316). Dykinson.

Gass, S. (1997). *Input, interaction, and the second language learner*. Lawrence Erlbaum.

Gilabert, R. (2005). Evaluating the use of multiple sources and methods in needs analysis: A case study of journalists in the autonomous community of Catalonia (Spain). In M. H. Long (Ed.), *Second language needs analysis* (pp. 182–99). Cambridge University Press.

Gilabert, R. (2007). Effects of manipulating task complexity on self-repairs during L2 production. *IRAL*, *45*, 215–240. https://doi.org/10.1515/iral.2007.010

Goldberg, A., Russell, M., & Cook, A. (2003). The effect of computers on student writing: A meta-analysis of studies from 1992 to 2002. *The Journal of Technology, Learning and Assessment*, *2*(1), 1-51

González-Cruz, B., Cerezo, L., & Nicolás-Conesa, F. (2022). A classroom-based study on the effects of WCF on accuracy in pen-and-paper versus computer-mediated collaborative writing. *Studies in Second Language Learning and Teaching*, *12*(4), 623-650. https://doi.org/10.14746/ssllt.2022.12.4.5

Harklau, L. (2002). The role of writing in classroom second language acquisition. *Journal of Second Language Writing*, *11*, 329–350. https://doi.org/10.1016/S1060-3743(02)00091-7

Hartshorn, K.J., Evans, N., Merrill, P., Sudweeks, R., Strong-Krause, D., & N. Anderson, N. (2010). Effects of dynamic corrective feedback on ESL writing accuracy. *TESOL Quarterly, 44*(1), 84-109. https://doi.org/10.5054/tq.2010.213781

Hockly, N. (2019). Automated writing evaluation. *ELT Journal*, 73(1), 82-88. https://doi.org/10.1093/elt/ccy044

Honeycutt, L. (2001). Comparing e-mail and synchronous conferencing in online peer response. *Written Communication*, *18*(1), 26–60. https://doi.org/10.1177%2F0741088301018001002

Isaias, P., Miranda, P. and Pifano, S., 2015. An empirical study on computer and paper based resources: are they competitive or complimentary means? *IADIS International Journal on Computer Science & Information Systems*, *10*(2), 129–144.

Ishikawa, T. (2007). The effect of increasing task complexity along the [Here-and-Now] dimension on L2 written narrative discourse. In M. P. Garcia Mayo (Ed.), *Investigating tasks in formal language learning* (pp. 136–156). Multilingual Matters.

Iwashita, N., & Dao, P. (2021). Peer Feedback in Second Language Oral Interaction. In H. Nassaji & E. Kartchava (Eds.), *The Cambridge Handbook of Corrective Feedback in Second Language Learning and Teaching* (pp. 275-299). Cambridge University Press. https://doi.org/10.1017/9781108589789.014

Johnson, M. D., Mercado, L., & Acevedo, A. (2012). The effect of planning sub-processes on L2 writing fluency, grammatical complexity, and lexical complexity. *Journal of Second Language Writing*, *21*(3), 264–282. https://doi.org/10.1016/j.jslw.2012.05.011

Kang, E. Y., & Han, Z. (2015). The efficacy of written corrective feedback in improving L2 written accuracy. *The Modern Language Journal*, *99*(1), 1–18. https://doi.org/10.1111/modl.12189

Kang, E. Y., & Han, Z. (2021). Written corrective feedback: Short-term and long-term effects on language learning. In R.M. Manchón & C. Polio (Eds.), *The Routledge handbook of second language acquisition and writing* (pp. 213-225). Routledge. https://doi.org/10.4324/9780429199691

Kepner, C. G. (1991) An experiment in the relationship of types of written feedback to the development of second-language writing skills. *Modern Language Journal, 75*(3), 305-313. https://doi.org/10.2307/328724

Kılıçkaya, F. (2019). Learners' perceptions of collaborative digital graphic writing based on semantic mapping. *Computer Assisted Language Learning*, *33*(1-2), 58-84. https://doi.org/10.1080/09588221.2018.1544912

Kim, S. (2010). Revising the revision process with Google Docs. In S. Kasten (Ed.), *TESOL classroom practice series. Effective second language writing* (pp. 171–177). TESOL Publications.

Kim, H. R., & Bowles, M. (2019). How Deeply Do Second Language Learners Process Written Corrective Feedback? Insights Gained from Think-Alouds. *Tesol Quarterly*, *53*(4), 913-938. https://doi.org/10.1002/tesq.522

Koltovskaia, S. (2020). Student engagement with automated written corrective feedback (AWCF) provided by Grammarly: A multiple case study. *Assessing Writing*, *44*, 100450. https://doi.org/10.1016/j.asw.2020.100450

Koltovskaia, S., & Mahapatra, S. (2022). Student engagement with computer-mediated teacher written corrective feedback: A case study. *Japan Association for Language Teaching Computer Assisted Language Learning Journal (JALT CALL Journal)*, *18*(2), 86–315. https://doi.org/10.29140/jaltcall.v18n2.519

Kormos, J. (2012). The role of individual differences in L2 writing. *Journal of second language writing*, *21*(4), 390-403. https://doi.org/10.1016/j.jslw.2012.09.003

Kormos, J. (In press/2023). The role of cognitive factors in second language writing and writing to learn a second language. *Studies in Second Language Acquisition.*

Kormos, J., & Trebits, A. (2012). The role of task complexity, modality, and aptitude in narrative task performance. *Language Learning*, *62*(2), 439-472. https://doi.org/10.1111/j.1467-9922.2012.00695.x

Lalande, J. F. (1982). Reducing composition errors: An experiment. *The Modern Language Journal*, *66*(2), 140-149. https://doi.org/10.2307/326382

Langer, J. A., and Applebee, A. N. (1987). *How writing shapes thinking: A study of teaching and learning*. NCTE Research Monograph 22. Urbana, IL.

Lavolette, E. H. (2014). *Effects of feedback timing and type on learning ESL grammar rules*. PhD Dissertation. Michigan State University.

Lavolette, E., Polio, C., & Kahng, J. (2015). The accuracy of computer-assisted feedback and students' responses to it. *Language Learning & Technology, 19*(2), 50-68

Lee, I. (2022). Developments in classroom-based research on L2 writing. *Studies in Second Language Learning and Teaching*, *12*(4), 551–574. https://doi.org/10.14746/ssllt.2022.12.4.2

Leow, R. P. (1997). Attention, awareness, and foreign language behavior. *Language Learning*, *47*, 467–506. https://doi.org/10.1111/0023-8333.00017

Leow, R. P. (2015). *Explicit learning in the L2 classroom: A student-centered approach*. Routledge. https://doi.org/10.4324/9781315887074

Leow, R. P. (2020). L2 writing-to-learn: Theory, research, and a curricular approach. In R.M. Manchón (Ed.), *Writing and language learning: Advancing research agendas* (pp. 95-117). John Benjamins. https://doi.org/10.1075/lllt.56.05leo

Leow, R. P., & Driver, M. (2021). Cognitive theoretical perspectives of corrective feedback. In H. Nassaji & E. Kartchava (Eds.), *The Cambridge handbook of corrective feedback in second language learning and teaching* (pp. 65-84). Cambridge University Press. https://doi.org/10.1017/9781108589789.004

Leow, R., P., & Manchón, R. M. (2022). Directions for future research agendas on L2 writing and feedback as language learning from an ISLA perspective. In R. M. Manchón & C. Polio (Eds.), *The Routledge handbook of second language acquisition and writing* (pp. 299-311). Routledge. https://doi.org/10.4324/9780429199691-31

Leow, R., & Suh, B-R. (2022). Theoretical perspectives on L2 writing, written corrective feedback, and language learning in individual writing conditions. In R. M. Manchón & C. Polio (Eds.), *The Routledge handbook of second language acquisition and writing* (pp. 9–21). Routledge.

Leow, R.P., Thinglum, A., & Leow, S.A. (2022). WCF processing in the L2 curriculum: A look at type of WCF, type of linguistic item, and L2 performance. *Studies in Second Language Learning and Teaching*, *12*(4), 651-673. https://doi.org/10.14746/ssllt.2022.12.4.6

Li, J., Link, S., & Hegelheimer, V. (2015). Rethinking the role of automated Writing evaluation (AWE) feedback in ESL writing instruction. *Journal of Second Language Writing*, *27*, 1–18. https://doi.org/10.1016/j.jslw.2014.10.004

Li, M. (2021). *Researching and teaching second language writing in the digital age*. Palgrave Macmillan.

Li, M., & Zhang, M. (2021). Collaborative writing in L2 classrooms: A research agenda. *Language Teaching*, *56*(1), 1-19.

Lin, Y. (2013). *The effects of task planning on L2 writing*. Unpublished Doctoral Dissertation, University of Auckland, New Zealand. Retrieved from https://researchspace.auckland.ac.nz/handle/2292/20926

Link, S., Mehrzad, M., & Rahimi, M. (2020). Impact of automated writing evaluation on teacher feedback, student revision, and writing improvement. *Computer Assisted Language Learning, 33*, 1–30. https://doi.org/10.1080/09588221.2020.1743323

Lira-Gonzales, M. L., Nassaji, H., & Chao Chao, K. W. (2021). Student engagement with teacher written corrective feedback in a French as a foreign language classroom. *Journal of Response to Writing*, *7*(2), 37-73.

Liu, Q., & Brown, D. (2015). Methodological synthesis of research on the effectiveness of corrective feedback in L2 writing. *Journal of Second Language Writing*, *30*, 66–8. https://doi.org/10.1016/j.jslw.2015.08.011

Liu, J., & Sadler, R. (2003). The effect and affect of peer review in electronic versus traditional modes on L2 writing. *Journal of English for Academic Purposes*, *2*(3), 193–227. https://doi.org/10.1016/S1475-1585(03)00025-0

López-Serrano, S., Roca de Larios, J.., & Manchón, R. M. (2020). Processing output during individual L2 writing tasks. In R. M. Manchón (Ed.), *Writing and Language Learning: Advancing research agendas*, (pp. 231-254). John Benjamins

Lu, X. (2010). Automatic analysis of syntactic complexity in second language writing *International Journal of Corpus Linguistics*, *15*, 474–496.

Lu, X. (2011). A corpus-based evaluation of syntactic complexity measures as indices of college-level ESL writers' language development. *TESOL Quarterly, 45*, 36–62. https://doi.org/10.5054/tq.2011.240859

Lu, X (2012). The relationship of rlexical richness to the quality of ESL learners' 0ral narratives. *The Modern Language Journal, 96*(2), 190-208.

Mackey, A., Gass, S., & McDonough, K. (2000). How do learners perceive interactional feedback? *Studies in Second Language Acquisition*, *22*(4), 471–497. https://doi.org/10.1017/S0272263100004010

Manchón, R. (Ed.). (2011a). *Learning-to-write and writing-to-learn in an additional language*. John Benjamins.

Manchón, R. M. (2011b). Writing to learn the language: Issues in theory and research. In R. M. Manchón (Ed.), *Learning-to-write and writing-to-learn in an additional language* (pp. 61-82). Johns Benjamins Company.

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

Manchón, R. M. (Ed.). (2020). *Writing and language learning: Advancing research agendas.* John Benjamins.

Manchón, R.M. (2023a). The psycholinguistics of second language writing. In A. Godfroid & H. Hopp (Eds.), *The handbook of second language acquisition and psycholinguistics* (pp. 400-412). Routledge

Manchón, R.M. (In press/2023b). Writing and language learning. In C. Chapelle (Ed.), *The Encyclopedia of Applied Linguistics*. Willey Blackwell

Manchón, R. & Cerezo, L. (2018). Writing as language learning. In J. Liontas (Ed.), *The TESOL encyclopedia of English language teaching*. John Wiley & Sons, Inc. https://doi.org/10.1002/9781118784235.eelt0530

Manchón, R. M., & Coyle, Y. (2022). Introduction to the special issue on L2 writing and feedback processing and use in pen and paper and digital environments: Advancing research and practice. *Studies in Second Language Learning and Teaching*, *12*(4), 541–550. https://doi.org/10.14746/ssllt.2022.12.4.1

Manchón, R.M., & Leow, R. P. (2020). Investigating the language learning potential of L2 writing. Methodological considerations for future research agendas. In R. M. Manchón (Ed.), *Writing and language learning. Advancing research agendas* (pp. 335-355). Benjamins.

Manchón, R. & Matsuda, P. (2016). *Handbook of second and foreign language writing*. De Gruyter Mouton. https://doi.org/10.1515/9781614511335

Manchón, R. M., Nicolás-Conesa, F., Cerezo, L., & Criado, R. (2020). L2 writers' processing of written corrective feedback. In W. Suzuki & N. Storch (Eds.), *Languaging in*

language learning and teaching: A collection of empirical studies (pp. 241-266). John Benjamins.

Manchón, R. M., & Roca de Larios, J. (2007). Writing-to-learn in instructed language learning contexts. In E. A. Soler & M. P. S. Jordá (Eds.), *Intercultural language use and language learning* (pp. 101–121). Springer. https://doi.org/10.1007/978-1-4020-5639-0_6

Manchón, R. M.., & Tardy, C. (Eds.). (2012) *Exploring L2 Writing-SLA interfaces* [Special Issue] *Journal of Second Language Writing*. https://doi.org/10.1016/j.jslw.2012.09.002

McBride, S. & Garcés-Manzanera, A. (2022). Higher Education L2 writing: a comparison of undergraduate perceptions on digital and pen-and-paper writing. In S. Flores Borjabad, A. Garcés-Manzanera & O. S. Ould García (Eds.), *El devenir de la lingüística y la cultura: Un estudio interdisciplinar sobre lengua, literatura y traducción* (pp. 1153-1171). Dykinson.

McBride, S., Manchón, R. (Forthcoming/2023). Analysing depth of processing in L2 writers' appropriation of written corrective feedback. Comparing the affordances of written languaging and think-aloud protocols. In R.M. Manchón & J. Roca de Larios (Eds), *Research methods in the study of L2 Writing Processes*. John Benjamins

Mohsen, M. A. (2022). Computer-mediated corrective feedback to improve L2 writing skills: A meta-analysis. *Journal of Educational Computing Research, 60*(5), *1-24.* https://doi.org/10.1177/07356331211064066

Moradian, M. R., Hossein-Nasabi, M., & Miriii, M. (2020). Effects of written languaging in response to direct and indirect corrective feedback on developing writing accuracy. In W. Suzuki & N. Storch (Eds.), *Languaging in language learning and teaching: A collection of empirical studies*, 55, 241 (pp. 267-286). John Benjamins.

Moradian, M. R., Miri, M., & Hossein Nasab, M. (2017). Contribution of written languaging to enhancing the efficiency of written corrective feedback. *International Journal of Applied Linguistics*, *27*(2), 406-426. https://doi.org/10.1111/ijal.12138

Murphy, L., & de Larios, J. R. (2010). Searching for words: One strategic use of the mother tongue by advanced Spanish EFL writers. *Journal of Second Language Writing*, *19*(2), 61-81.

Nabei, T. and Swain, M. 2002. Learner awareness of recasts in classroom interaction: A case study of an adult EFL student's second language learning. *Language Awareness*, *11*(1): 43–63. https://doi.org/10.1080/09658410208667045

Nassaji, H., & Kartchava, E. (2021). Corrective feedback in second language teaching and learning. In H. Nassaji & E. Kartchava (Eds.), *The Cambridge handbook of corrective feedback in second language learning and teaching* (pp. 1-20). Cambridge University Press. https://doi.org/10.1017/9781108589789.001

Newell, G.E. (1984). Learning from writing in two content areas: A case study/protocol analysis. *Research in the Teaching of English*, *18*(3), 265-287.

Nicolás–Conesa, F., Manchón, R. M., & Cerezo, L. (2019). The effect of unfocused direct and indirect written corrective feedback on rewritten texts and new texts: Looking into feedback for accuracy and feedback for acquisition. *The Modern Language Journal*, *103*(4), 848-873. https://doi.org/10.1111/modl.12592

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). John Benjamins. https://doi.org/10.1075/tblt.7.05nit

Nurmukhamedov, U. (2009). Teacher feedback on writing: Considering the options. *Writing & Pedagogy*, *1*(1), 113-124. https://doi.org/10.1558/wap.v1i1.113

Ortega, L. (2003). Syntactic complexity measures and their relationship to L2 proficiency: A research synthesis of college-level L2 writing. *Applied Linguistics*, *24* (4), 492-518.

Ortega, L. (2009). Interaction and attention to form in L2 text-based computer-mediated communication. In A. Mackey & C. Polio (Eds.), *Multiple perspectives on interaction* (pp. 226-253). Routledge.

Park, E. S., & Kim, O. Y. (2019). Learners' engagement with indirect written corrective feedback: Depth of processing and self-correction. In R.P. Leow (Ed.), *The Routledge handbook of second language research in classroom learning* (pp. 212-226). Routledge.

Plonsky L, Sudina E, Hu Y (2021). Applying meta-analysis to research on bilingualism: An introduction. *Bilingualism: Language and Cognition, 24*(5), 19 – 824. https://doi.org/10.1017/S1366728920000760

Polio, C. (2012). The relevance of second language acquisition theory to the written error correction debate. *Journal of Second Language Writing, 21*(4), 375-389.

Polio, C., Fleck, C. and Leder, N. (1998) 'If only I had more time': ESL learners' changes in linguistic accuracy on essay revisions. *Journal of Second Language Writing*, *7*,43–68. https://doi.org/10.1016/S1060-3743(98)90005-4

Polio, C., & Lee, J. (2017). Written language learning. In S. Loewen & M. Sato (Eds.), *The Routledge handbook of instructed second language acquisition* (pp. 299-317). Routledge.

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*(4), 277-303. https://doi.org/10.1016/S1060-3743(01)00046-7

Raimes, A. (1978). Language proficiency, writing ability, and composing strategies: A study of ESL college student writers. *Language Learning*, *37*(3), 439–468. https://doi.org/10.1111/j.1467-1770.1987.tb00579.x

Révész, A., Michel, M., & Gilabert, R. (2016). Measuring cognitive task demands using dual-task methodology, subjective self-ratings, and expert judgments: A validation study. *Studies in Second Language Acquisition, 38*(4), 703-737. https://doi.org/10.1017/S0272263115000339

Robb, T., Ross, S., & Shortreed, I. (1986). Salience of feedback on error and its effect on EFL writing quality. *Tesol Quarterly*, *20*(1), 83-96. https://doi.org/10.2307/3586390

Robinson, P. (2001). Task complexity, task difficulty, and task production: Exploring interactions in a componential framework. *Applied Linguistics*, *22*(1), 27-57. https://doi.org/10.1093/applin/22.1.27

Robinson, P. (Ed.). (2011). *Second language task complexity: Researching the cognition hypothesis of language learning and performance* John Benjamins.

Roca de Larios, J. R., & Coyle, Y. (2021). Learners' Engagement with Written Corrective Feedback in Individual and Collaborative L2 Writing Conditions. In R.M. Manchón & C. Polio. (Eds.), *The Routledge Handbook of Second Language Acquisition and Writing* (pp. 81-93). Routledge.

Rummel, S., & Bitchener, J. (2015). The effectiveness of written corrective feedback and the impact Lao learners' beliefs have on uptake. *Australian Review of Applied Linguistics*, *38*(1), 66-84. https://doi.org/10.1075/aral.38.1.04rum

Sachs, R., & Polio, C. (2007). Learners' uses of two types of written feedback on a l2 writing revision task. *Studies in Second Language Acquisition*, *29*(1), 67-100. https://doi.org/10.1017/S0272263107070039

Salazar, P. (2012). Exploring learners' noticing of corrective feedback through stimulated recall. In E. Alcón-Soler & M. P Safont-Jordá (Eds.), *Discourse and language learning across L2 instructional settings* (pp. 279-293). Brill. https://doi.org/10.1163/9789401208598_014

Sánchez, A. J., Manchón, R. M., & Gilabert, R. (2020). The effects of task repetition across modalities and proficiency levels. In R.M. Manchón (Ed.), *Writing and language learning: Advancing research agendas* (pp. 121-144). John Benjamins

Saragih, N, A., Madya, S., Siregar, R.A., Saragih, W. (2021). Written corrective feedback: students' perception and preferences. *International Online Journal of Education and Teaching (IOJET), 8*(2). 676-690.

Sarre, C., Grosbois, M., & Brudermann, C. (2019). Fostering accuracy in L2 writing: Impact of different types of corrective feedback in an experimental blended learning EFL course. *Computer Assisted Language Learning, 31*, 1–23. https://doi.org/10.1080/09588221.2019. 1635164

Schmidt, R. (1990). The role of consciousness in second language learning. *Applied Linguistics*, *11*(2), 129-158. https://doi.org/10.1093/applin/11.2.129

Schmidt, R. (1995). Consciousness and foreign language learning: A tutorial on the role of attention and awareness in learning. In R. Schmidt (Ed.), *Attention and awareness in foreign language learning*. Honolulu, HI: University of Hawaii, Second Language Teaching and Curriculum Center, 1–63.

Schmidt, R. (2001). Attention. In P. Robinson (Ed.), *Cognition and second language instruction* (pp. 3-32). Cambridge University Press

Schultz, J. M., Warschauer, M., & Kern, R. (2000). Computers and collaborative writing. In M. Warschauer & R. Kern (Eds.), *Network-based language teaching: Concepts and practice* (pp. 121-150), Cambridge University Press.

Sheen, Y. (2007) The effect of focused written corrective feedback and language aptitude on ESL learners' acquisition of articles. *TESOL Quarterly, 41*,255–83. https://doi.org/10.1002/j.1545-7249.2007.tb00059.x

Sheen, Y., & Lyster, R. (Eds.). (2010) *The role of oral and written corrective feedback in SLA* [Special Issue] *Studies in Second Language Acquisition*. https://doi.org/10.1017/S0272263109990489

Sheen, Y., Wright, D., & Moldawa, A. (2009). Differential effects of focused and unfocused written correction on the accurate use of grammatical forms by adult ESL learners. *System, 37*(4), 556-569. https://doi.org/10.1016/j.system.2009.09.002

Sheppard, K. (1992). Two feedback types: Do they make a difference?. *RELC journal*, *23*(1), 103-110. https://doi.org/10.1177%2F003368829202300107

Shintani, N. (2015). The effects of computer-mediated synchronous and asynchronous direct corrective feedback on writing: A case study. *Computer Assisted Language Learning*,

29(3), 517-538. https://doi.org/10.1080/09588221.2014.993400

Shintani, N., & Aubrey, S. (2016). The effectiveness of synchronous and Asynchronous written corrective feedback on grammatical accuracy in a computer-mediated environment. *The Modern Language Journal*, *100*(1), 296-319. https://doi.org/10.1111/modl.12317

Shintani, N., & Ellis, R. (2013). The comparative effect of direct written corrective feedback and metalinguistic explanation on learners' explicit and implicit knowledge of the English indefinite article. *Journal of Second Language Writing*, *22*(3), 286-306. https://doi.org/10.1016/j.jslw.2013.03.011

Shintani, N., Ellis, R. and Suzuki, W. (2014) Effects of written feedback and revision on learners' accuracy in using two English grammatical structures. *Language Learning,* 64, 103–131. https://doi.org/10.1111/lang.12029

Simard, D., & Zuniga, M. (2020). Exploring the mediating role of emotions expressed in L2 written languaging in ESL learner text revisions. In W. Suzuki & N. Stroch (Eds.), *Languaging in language learning and teaching: A collection of empirical studies* (pp. 287-307). John Benjamins.

Skehan, P. (2009). Modelling second language performance: Integrating complexity, accuracy, fluency, and lexis. *Applied Linguistics*, *30*, 510-532.

Sommers, N. (1982). Responding to student writing. *College composition and communication*, *33*(2), 148-156. https://doi.org/10.2307/357622

Stevenson, M., & Phakiti, A. (2014). The effects of computer-generated feedback on the quality of writing. *Assessing Writing, 19*, 51–65. https://doi.org/10.1016/j.asw.2013.11.007

Storch, N. (2013). *Collaborative writing in L2 classrooms*. Multilingual Matters. https://doi.org/10.21832/9781847699954

Storch, N. (2018). Written corrective feedback from sociocultural theoretical perspectives: A research agenda. *Language Teaching*, *51*(2), 262-277.

Storch, N., & Wigglesworth, G. (2010). Learners' processing, uptake, and retention of corrective feedback on writing: Case studies. *Studies in Second Language Acquisition*, *32*(2), 303-334. https://doi.org/10.1017/S0272263109990532

Suh, B.-R. (2020). Are think-alouds reactive? Evidence from an L2 written corrective feedback study. *Language Teaching Research*, 1- 21. https://doi.org/10.1177/1362168820967166

Suzuki, W. (2012). Written languaging, direct correction, and second language writing

revision. *Language Learning*, *62*, 1110-1133. https://doi.org/10.1111/j.1467-9922.2012.00720.x

Suzuki, W. (2017). The effect of quality of written languaging on second language learning. *Writing & Pedagogy*, *8*(3), 461-482. https://doi.org/10.1558/wap.27291

Suzuki, W., Ishikawa, M., & Storch, N. (Forthcoming/2023). Verbally mediated data: Written languaging. In R.M. Manchón & J. Roca de Larios (Eds), *Research methods in the study of L2 writing processes*. John Benjamins

Suzuki, W., & Storch, N. (Eds.). (2020). *Languaging in language learning and teaching: A collection of empirical studies*. John Benjamins.

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). Newbury House.

Swain, M. (2000). The output hypothesis and beyond: mediating acquisition through collaborative dialogue. In J.P. Lantolf, (Ed.), *Sociocultural theory and second language learning* (pp. 97-114). Oxford University Press.

Swain, M. (2005). The output hypothesis: Theory and research. In E. Hinkel (Ed.), *Handbook of research in second language teaching and learning* (pp. 495-508). Routledge.

Swain M. (2006). Languaging, agency and collaboration in advanced second language proficiency. In Byrnes H. (Ed.), *Advanced language learning: The contribution of Halliday and Vygotsky* (pp.95-108). Continuum.

Swain, M., & Lapkin, S. (2002). Talking it through: Two French immersion learners' response to reformulation. *International Journal of Educational Research*, *37*(3-4), 285-304. https://doi.org/10.1016/S0883-0355(03)00006-5

Swain, M., & Watanabe, Y. (2013). Languaging: Collaborative dialogue as a source of second language learning. In C. Chapelle (Ed.), *The encyclopaedia of applied linguistics* (pp. 3218-3225). Willey Blackwell. http://doi.org/10.1002/9781405198431.wbeal0664

Tafazoli, D., Nosratzadeh, H., & Hosseini, N. (2014). Computer-mediated corrective feedback in ESP courses: Reducing grammatical errors via email. *Procedia-Social and Behavioral Sciences*, *136*, 355 -359. https://doi.org/10.1016/j.sbspro.2014.05.341

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

Tharp, R. G., & Gallimore, R. (1991). *Rousing minds to life: Teaching, learning, and schooling in social context*. Cambridge University Press.

Truscott, J. (1996). The case against grammar correction in L2 writing classes. *Language learning*, *46*(2), 327-369. https://doi.org/10.1111/j.1467-1770.1996.tb01238.x

Truscott, J. and Hsu, A. Y. (2008) Error correction, revision, and learning. *Journal of Second Language Writing 17*,292–305. https://doi.org/10.1016/j.jslw.2008.05.003

Tuzi, F. (2004). The impact of e-feedback on the revisions of L2 writers in an academic writing course. *Computers and composition*, *21*(2), 217-235. https://doi.org/10.1016/j.compcom.2004.02.003

Van Beuningen, C. G., De Jong, N. H., & Kuiken, F. (2008). The effect of direct and indirect corrective feedback on L2 learner's written accuracy. *International Journal of Applied Linguistics*. *156*, 279-296. https://doi.org/10.2143/ITL.156.0.2034439

Van Beuningen C. G., De Jong, N. H. and Kuiken, F. (2012) Evidence on the effectiveness of comprehensive error correction in second language writing. *Language Learning,* 62, 1–41. https://doi.org/10.1111/j.1467-9922.2011.00674.x

Vasylets, O., & Gilabert, R. (2021). Task effects across modalities. In R. M. Manchón & C. Polio (Eds.). *The Routledge handbook of second language acquisition and writing* (pp. 39-51). Routledge

Vasylets, O., Gilabert, R., & Manchón, R. M. (2017). The effects of mode and task complexity on second language production. *Language Learning*, *67*(2), 394-430. https://doi.org/10.1111/lang.12228

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. https://doi.org/10.1558/isla.38289

Vasylets, O., Gilabert, R., & Manchón, R.M. (2020). Task modality and task complexity. In R. M. Manchón (Ed.), *Writing and language learning. Advancing research agendas*, (pp. 183-206). John Benjamins

Vasylets, O., Mellado, M.D., Plonsky, L. (2022). The role of cognitive individual differences in digital versus pen-and-paper writing. *Studies in Second Language Learning and Teaching*, *12*(4), 721-743. https://doi.org/10.14746/ssllt.2022.12.4.9

Vincent, J. (2016). Students' use of paper and pen versus digital media in university environments for writing and reading–a cross-cultural exploration. *Journal of Print*

Media and Media Technology Research, *5*(2), 97-106. https://doi.org/10.14622/JPMTR-1602

Vu, D. V., & Peters, E. (2020). Learning vocabulary from reading-only, reading- while-listening, and reading with textual input enhancement: Insights from Vietnamese EFL learners. *RELC Journal*, 53(1), 1-16. https:// doi.org/10.1177/0033688220911485

Vu, D. V., & Peters, E. (2023). A longitudinal study on the effect of mode of reading on incidental collocation learning and predictors of learning gains. *TESOL Quarterly*, 57(1), 5-32. https://doi.org/10.1002/tesq.3111

Vygotsky, L. (1978). Interaction between learning and development. In *Mind and society* (pp. 79-91). Cambridge, MA: Harvard University Press.

Ware, P. (2004). Confidence and competition online: ESL student perspectives on web-based discussions in the classroom. *Computers and Composition*, *21*(4), 451-468. http://doi.org/10.1016/j.compcom.2004.08.004.

Ware, P. D., & O'Dowd, R. (2008). Peer feedback on language form in telecollaboration. *Language Learning and Technology*, *12*(1), 43–63. http://doi.org/10125/44130

Whithaus, C., Harrison, S. B., & Midyette, J. (2008). Keyboarding compared with handwriting on a high-stakes writing assessment: Student choice of composing medium, raters' perceptions, and text quality. *Assessing Writing, 13*, 4–25. https://doi.org/10.1016/j.asw.2008.03.001

Wigglesworth, G., & Storch, N. (2012a). Feedback and writing development through collaboration: A socio-cultural approach. In R. M. Manchón (Ed*.), L2 writing development: Multiple perspectives* (pp. 69-99). Mouton de Gruyter. https://doi.org/10.1515/9781934078303

Wigglesworth, G., & Storch, N. (2012b). What role for collaboration in writing and writing feedback. *Journal of Second Language Writing*, *21*, 364-374. https://doi.org/10.1016/j.jslw.2012.09.005

Wolfe, E. W., Bolton, S., Feltovich, B., & Niday, D. M. (1996). The influence of student experience with word processors on the quality of essays written for a direct writing assessment. *Assessing Writing, 3*, 123–147. https://doi.org/10.1016/S1075-2935(96)90010-0

Wolfe–Quintero, K., Inagaki, S., & Kim, H.-Y. (1998). *Second language development in writing: Measures of fluency, accuracy, and complexity*. Honolulu, HI: Second Language Teaching & Curriculum Center, University of Hawai'i at Manoa.

Wood, D., Bruner, J., & Ross, G. (1976). The role of tutoring in problem-solving. *Journal of*

Child Psychology and Psychiatry and Allied Disciplines, *17*, 89100. https://doi.org/10.1111/j.1469-7610.1976.tb00381.x

Yang, Y. (2010). Students' reflection on online self-correction and peer review to improve writing. *Computer and Education, 55*, 1202- 1210

Yang, C., Hu, G., & Zhang, L. J. (2014). Reactivity of concurrent verbal reporting in second language writing. *Journal of Second Language Writing*, *24*(1), 51–70. https://doi.org/10.1016/j.jslw.2014.03.002

Yang, C., Zhang, L. J., & Parr, J. M. (2020). The reactivity of think-alouds in writing research: Quantitative and qualitative evidence from writing in English as a foreign language. *Reading and Writing*, *33*(2), 451-483. https://doi.org/10.1007/s11145-019-09970-7

Yoon, H., & Polio, C. (2017). The linguistic development of students of English as a second language in two written genres. *TESOL Quarterly*, *51*, 275–301. https://doi.org/10.1002/tesq.296

Zalbidea, J. (2017). 'One task fits all'? The roles of task complexity, modality, and working memory capacity in L2 performance. *The Modern Language Journal*, *101*(2), 335-352. https://doi.org/10.1111/modl.12389

Zhang, L. J., & Zhang, D. (2019). Think-aloud protocols. In J. McKinley & H. Rose (Eds.), *The Routledge handbook of research methods in applied linguistics* (pp. 302-311). Routledge.

Zhang, Z. V. (2020). Engaging with automated writing evaluation (AWE) feedback on L2 writing: Student perceptions and revisions. *Assessing Writing, 43,* 100439. https://doi.org/10.1016/j.asw.2019.100439.

Zhang, Z.V., & Hyland, K. (2018). Student engagement with teacher and automated feedback on L2 writing. *Assessing Writing*, *36*, 90–102. https://doi.org/10.1016/j.asw.2018.02.004

Zhi, M., & Huang, B. (2021). Investigating the authenticity of computer-and paper-based ESL writing tests. *Assessing Writing* 50. https://doi.org/10.1016/j.asw.2021.100548

APPENDICES

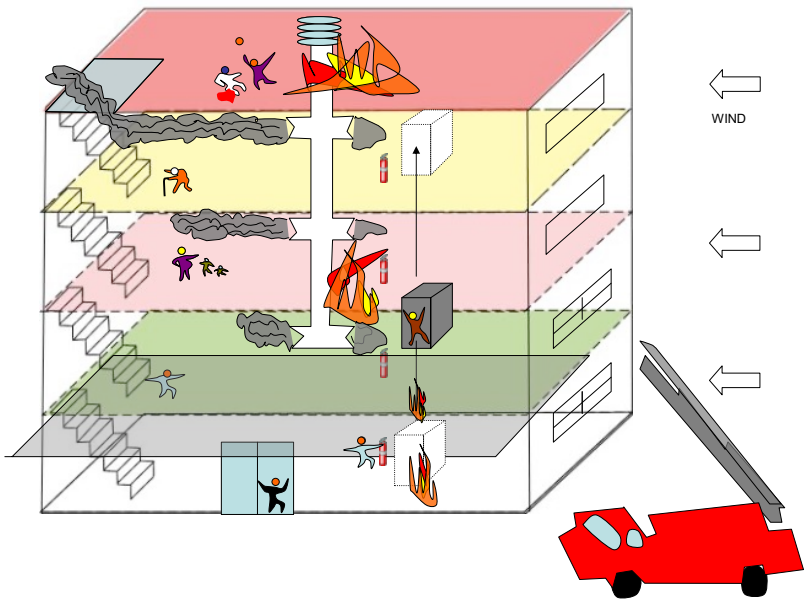
APPENDIX 1

Name:
Age:
Date:

You begin writing at:
You finish writing at:

Topic: Fire Chief

Instructions: Observe the fire chief image and **write a description**, considering all of the different elements involved in the situation. More precisely, you have to explain (a) which action you would take to save as many people as possible, and (b) the sequence (i.e., the order) in which you would take those actions. In both cases, you have to justify your choice of actions and their sequence. [In short, **say what you would do, in which order, and why**.]



APPENDIX 2

UNIVERSIDAD DE MURCIA

Name: _____ Age ___
Date: ____
You begin writing at _____ h _____ min
You finish writing at _____ h _____ min

PROCESAMIENTO DEL FEEDBACK DIRECTO

Instructions: Copy into the first column (Error) each of the errors identified in your text. Next, copy the correction of each error into the second column (Correction). Following this, write the code that corresponds to each error in the third column (Code). Finally, provide an explanation for each error made in the fourth column (Explanation).

Codes:
L for lexis
GR for Grammar
SP for spelling
P for punctuation
O for other

	ERROR	CORRECTION	CODE	EXPLANATION
1				
2				
3				
4				

APPENDIX 3

Name: _____ Age ___
Date: ____
You begin writing at _____ h _____ min
You finish writing at _____ h _____ min

PROCESAMIENTO DEL FEEDBACK DIRECTO

Instructions: In this task I am interested in hearing what you think about when completing the error form below. So, I am going to ask you to speak aloud and say anything and everything that crosses your mind the entire time you are working. I would like you to talk constantly, without planning what you are going to say. Imagine you are in an empty room, talking to yourself. To complete the task, first, copy into the first column (Error) each of the errors identified in your text. Next, copy the correction of each error into the second column (Correction) and the code (using the error-coding sheet) for the error in the third column (Code). Finally, provide an explanation for each error made in the last column (Explanation). Remember to keep talking throughout and just say whatever you are thinking. You have a maximum of one hour to complete the task.

	ERROR	CORRECTION	CODE	EXPLANATION
1				
2				
3				

APPENDIX 4

UNIVERSIDAD DE MURCIA

Name

Age:

Date:

Final Questionnaire

1. Did you find the feedback provided (that is, the corrections provided in red above the mistakes you made in the first version of the text) useful for making fewer errors in the final version of the text? Please answer yes or no and explain why. (NOTE: If you need more space to answer the question, use the back of this sheet.)

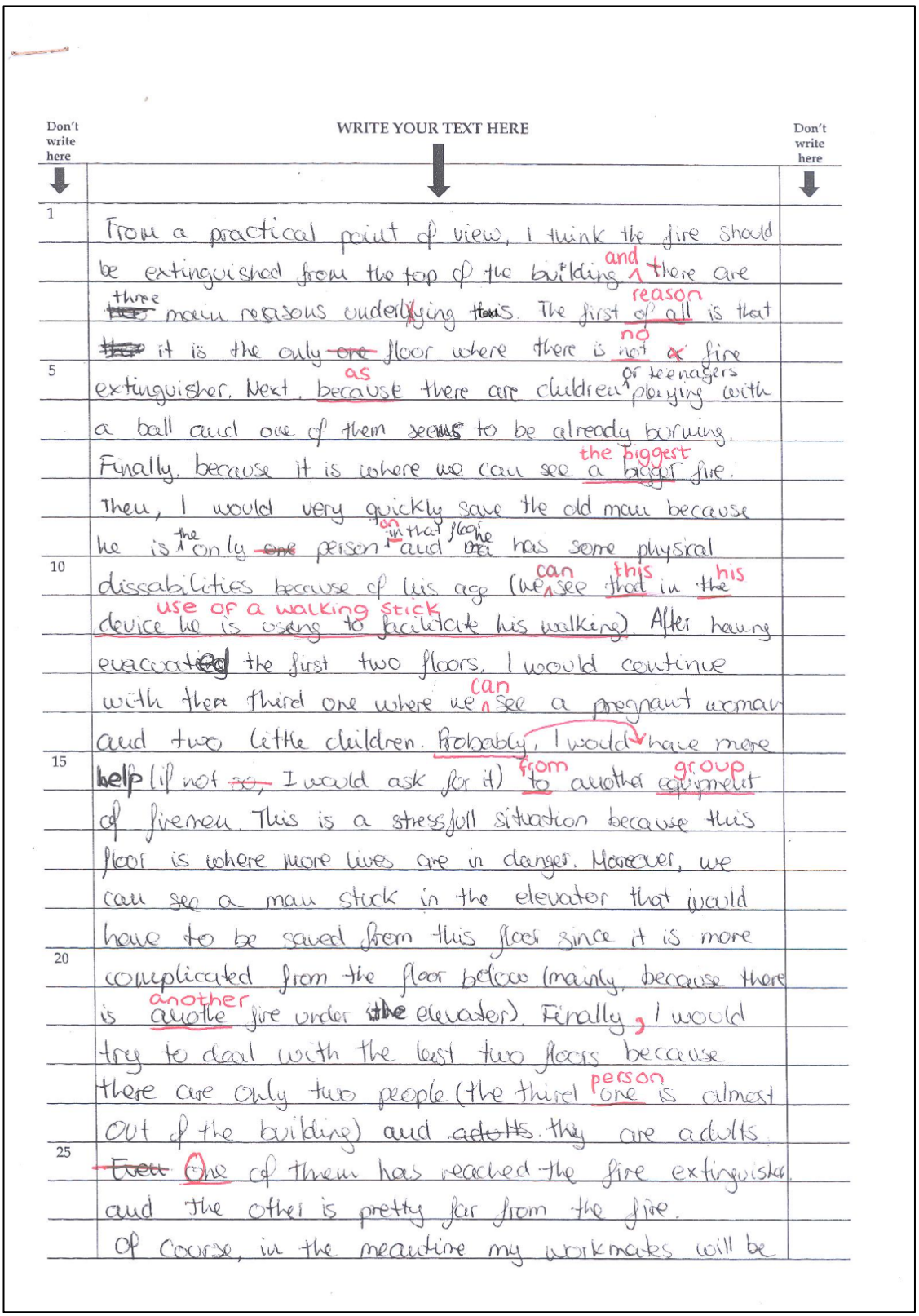
2. When re-writing the text, did you incorporate all of the corrections we provided you with? Mark an option with X.
(NOTE: If you need more space to answer the question, use the back of this sheet.)

Sí: ____

No: ____

3. If you DID NOT incorporate all of the corrections provided, indicate, for each of the errors you did not incorporate, if it concerns a grammatical, vocabulary or spelling error. Add any reasoning that you believe necessary.
(NOTE: If you need more space to answer the question, use the back of this sheet.)

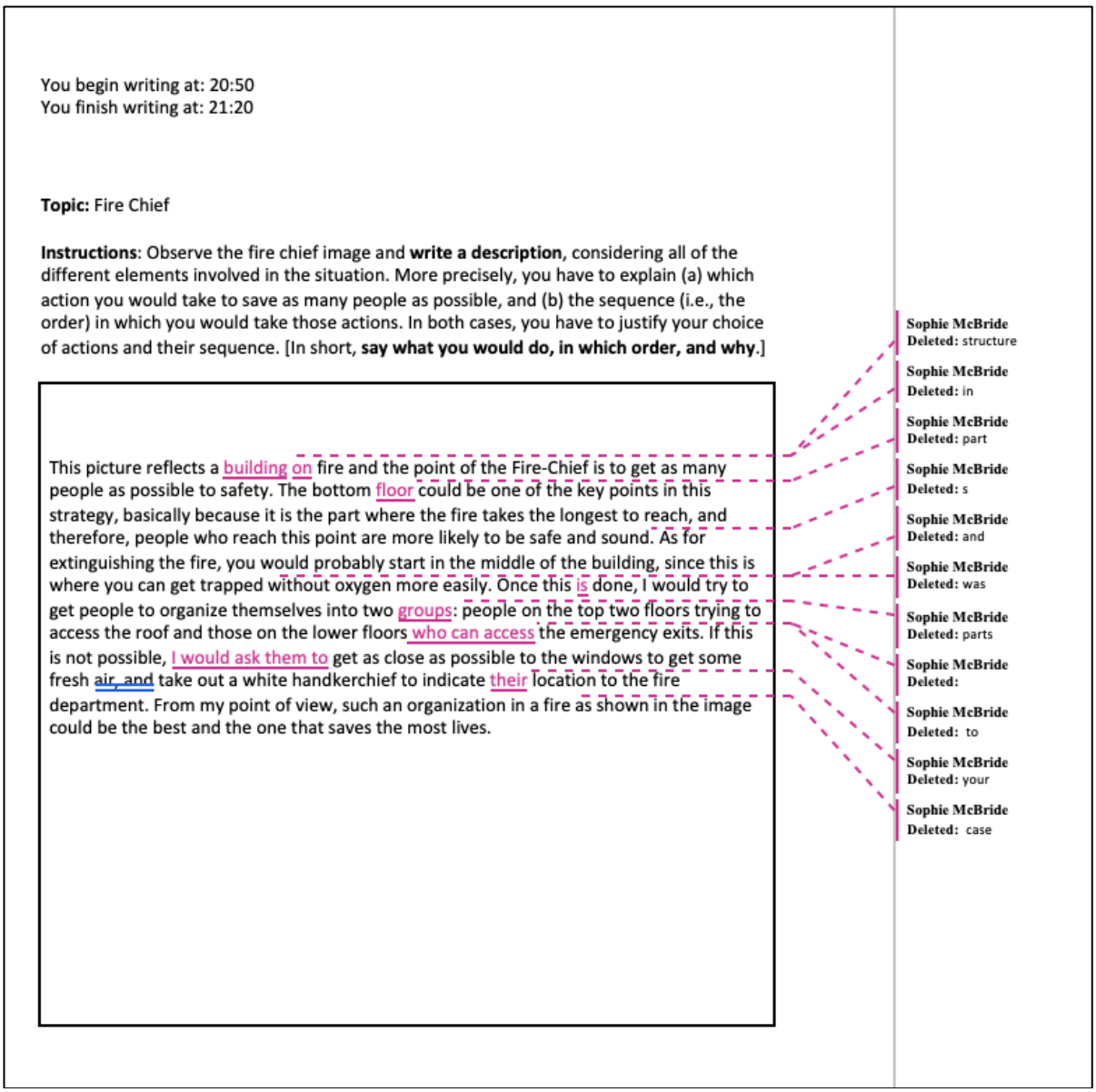
APPENDIX 5



Don't write here | WRITE YOUR TEXT HERE | Don't write here

From a practical point of view, I think the fire should be extinguished from the top of the building and there are three main reasons underlying this. The first reason is that it is the only floor where there is no fire extinguisher. Next, as there are children or teenagers playing with a ball and one of them seems to be already boring. Finally, because it is where we can see the biggest fire. Then, I would very quickly save the old man because he is the only person on that floor and he has some physical dissabilities because of his age (we can see this in the use of a walking stick device he is using to facilitate his walking). After having evacuated the first two floors, I would continue with the third one where we can see a pregnant woman and two little children. Probably, I would have more help (if not, I would ask for it) from another group of firemen. This is a stressfull situation because this floor is where more lives are in danger. Moreover, we can see a man stuck in the elevator that would have to be saved from this floor since it is more complicated from the floor below (mainly, because there is another fire under the elevator). Finally, I would try to deal with the last two floors because there are only two people (the third person is almost out of the building) and they are adults. One of them has reached the fire extinguisher and the other is pretty far from the fire. Of course, in the meantime my workmates will be

APPENDIX 6



You begin writing at: 20:50
You finish writing at: 21:20

Topic: Fire Chief

Instructions: Observe the fire chief image and **write a description**, considering all of the different elements involved in the situation. More precisely, you have to explain (a) which action you would take to save as many people as possible, and (b) the sequence (i.e., the order) in which you would take those actions. In both cases, you have to justify your choice of actions and their sequence. [In short, **say what you would do, in which order, and why**.]

This picture reflects a building on fire and the point of the Fire-Chief is to get as many people as possible to safety. The bottom floor could be one of the key points in this strategy, basically because it is the part where the fire takes the longest to reach, and therefore, people who reach this point are more likely to be safe and sound. As for extinguishing the fire, you would probably start in the middle of the building, since this is where you can get trapped without oxygen more easily. Once this is done, I would try to get people to organize themselves into two groups: people on the top two floors trying to access the roof and those on the lower floors who can access the emergency exits. If this is not possible, I would ask them to get as close as possible to the windows to get some fresh air, and take out a white handkerchief to indicate their location to the fire department. From my point of view, such an organization in a fire as shown in the image could be the best and the one that saves the most lives.

Sophie McBride
Deleted: structure

Sophie McBride
Deleted: in

Sophie McBride
Deleted: part

Sophie McBride
Deleted: s

Sophie McBride
Deleted: and

Sophie McBride
Deleted: was

Sophie McBride
Deleted: parts

Sophie McBride
Deleted:

Sophie McBride
Deleted: to

Sophie McBride
Deleted: your

Sophie McBride
Deleted: case