Examining two form-focused instruction combinations and their relationship with language-analytic ability in young learners: A study on the English possessive determiners *his/her*

María Martínez-Adrián
Department of English, German and Translation and Interpretation Studies
University of the Basque Country UPV/EHU
maria.martineza@ehu.eus

Kevin Iglesias-Diéguez
Department of English, German and Translation and Interpretation Studies
University of the Basque Country UPV/EHU
kevin.iglesias@ehu.eus

Abstract

Recent work conducted in input-limited classrooms with young learners has made a call for the incorporation of form-focused instruction (FFI) in the form of focused tasks and/or additional treatments that could help learners expand their attention to grammar. Current research has also uncovered the need for the examination of the interaction between FFI, individual differences, such as language-analytic ability, and language achievement. Thus, this paper examines the effect of two forms of FFI (dictogloss+self-correction and dictogloss+metalinguistic explanations [ME]) on learners' accuracy in the use of the possessive determiners *his/her* in the final product of the dictogloss tasks performed. It also looks into the interface between FFI and language-analytic ability as measured by the MLAT-ES for child language. To do so, 33 Basque/Spanish bilingual learners of English (aged 10-11) were split into a dictogloss+self-correction and a dictogloss+ME group. Participants were asked to complete three collaborative dictogloss tasks targeting the possessive determiners *his/her* in three consecutive weeks. The dictogloss+ME group achieved greater accuracy in the use of the possessive determiners *his/her* than the group with self-correction. Likewise, the benefits observed in the former were independent of participants' language-analytic ability, all of which suggests a blurring effect of the ME treatment received.

Keywords: form-focused instruction (FFI); metalinguistic explanations (ME); language-analytic ability; young learners of English as a Foreign Language (EFL); dictogloss.

Resumen

Trabajos recientes llevados a cabo con jóvenes aprendices en contextos de input limitado señalan la necesidad de la incorporación de instrucción centrada en la forma (FFI) mediante tareas focalizadas y/o tratamientos adicionales que podrían ayudarles a prestar mayor atención a la gramática. Investigaciones actuales también han enfatizado la necesidad de examinar la interacción entre FFI, las diferencias individuales, como la habilidad analítica sobre la lengua, y las ganancias lingüísticas. Por tanto, este artículo analiza el efecto de dos tipos de FFI (dictoglosia+auto-corrección y dictoglosia+explicaciones metalingüísticas [ME]) sobre la precisión en el uso de los determinantes posesivos en inglés *his/her* en el producto final de la dictoglosia. También se centra en la interfaz entre FFI y la habilidad analítica sobre la lengua medida a través del MLAT-ES, diseñado para jóvenes aprendices. Para ello, se dividió a 33 aprendices de inglés bilingües en euskera y español (de 10 y 11 años de edad) en dos grupos: dictoglosia+auto-corrección y dictoglosia+ME. Los participantes completaron tres dictoglosias centradas en los determinantes posesivos *his/her* en tres semanas no consecutivas. El grupo dictoglosia+ME obtuvo un incremento mayor en su precisión en el uso de los determinantes posesivos *his/her* que el grupo dictoglosia+auto-corrección. Además, los beneficios observados fueron independientes de la habilidad analítica sobre la lengua de los participantes, lo cual sugiere que el tratamiento empleado puede beneficiar a todos los participantes por igual.

Palabras clave: instrucción centrada en la forma; explicaciones metalingüísticas; habilidad analítica sobre la lengua; jóvenes aprendices de inglés como lengua extranjera; dictoglosia.

1. Introduction

The investigation of attention to form operationalised by language-related episodes (LREs) during the performance of collaborative tasks in the case of young learners has attracted the attention of researchers in the last decade. The existing differences between adults and children in how they approach the language learning process makes research with young learners a priority as previous research findings coming from adult populations cannot be overextended to young learners (Oliver & Azkarai, 2017). In this respect, recent investigations conducted with young learners while performing collaborative speaking+writing tasks have revealed that, unlike adults, LREs are not so elaborated and meaning- and form-related aspects receive

attention in equal terms (Martínez-Adrián & Gallardo-del-Puerto, 2021). In the light of the results obtained, there is an acute need to expand learners' attention to grammar forms through focused tasks alone or in combination with additional treatments, which are techniques encompassed by form-focused instruction (FFI). Even if this line of research is receiving increasing attention in the last years in input-limited settings (Calzada, 2021; Gorman & Ellis, 2019; Serrano, 2011), it is still in its infancy, and more investigations targeting different types of FFI are specially needed.

In particular, even though previous research has attested the benefits of focused tasks such as dictogloss tasks in drawing learners' attention to formal features, these seem to depend on the specific target, this task offering greater advantages in the case of the third person singular *-s* morpheme than in the case of articles (Calzada & García Mayo, 2020). Likewise, some additional treatments to the implementation of dictogloss tasks have not yielded the expected results (Calzada, 2021). Thus, the effectiveness of the incorporation of FFI in the form of focused tasks that elicit specific linguistic features and additional treatments that could maximise the potential of these collaborative tasks by raising learners' awareness of these features either before or after their performance (Loewen & Sato, 2018) merits further investigation in young learners.

A call in recent research has also been made as regards the interaction between FFI and individual differences (IDs) (Roehr-Brackin & Tellier, 2019; Suzuki, 2022; Suzuki & Dekeyser, 2017). Among these IDs, language-analytic ability has been suggested as a predictor of success for the acquisition of grammar properties (Li, 2016; Roehr-Brackin & Tellier, 2019) and FFI to have the capacity to level out these differences (Suzuki, 2022). However, research along these lines is scarce, especially with young learners. In particular, how different forms of FFI may affect the relationship between language-analytic ability and grammar accuracy remains unknown.

This paper aims to fill these gaps by comparing the written production of two groups of Basque/Spanish bilingual young learners of English who have received different types of FFI. Partially building on Gorman and Ellis (2019), one group performed three collaborative focused tasks (i.e. a dictogloss task) and was allowed to self-correct their written production before starting the following task, whereas the other group received two sessions of metalinguistic explanations (ME) in addition to the collaborative dictogloss tasks with no self-correction.

Examining their gains in accuracy in the use of the English possessive determiners ***his/her*** can shed light on the effectiveness of different types and combinations of FFI on language production. More specifically, a collaborative task such as a dictogloss, unspeeded because of its writing component, generates reflection about language, and can force learners to draw upon any metalinguistic knowledge they might have (Ellis, 2003;

Izumi, 2003; Kuiken & Vedder, 2012).[1] In particular, ME can draw students' attention to possessive determiners and trigger its noticing prior to the start of a new task, helping them to build their metalinguistic representations. During the performance of these tasks, learners might resort to the explicit knowledge about these features developed during the ME sessions either in an overt or covert way (Gutiérrez, 2013). As a result, the availability of metalinguistic representations can lead to an increase in learners' accuracy of language forms during production (Johns, 2003; Mitchell, 2000; Thepseenu & Roehr, 2013) and an impact on language development (Ellis, 2003; Erlam, 2013; Izumi, 2003). Moreover, an analysis of how their gains vary with respect to their language-analytic ability can provide further insight into the interplay between FFI and students' IDs.

This paper is structured as follows. Section 2 will review the main findings concerning FFI interventions and highlight the relevance of language-analytic ability in language learning. In sections 3 and 4, the research questions will be posed, and the methodology followed to answer them will be presented. Section 5 will describe the results of the study, which we will discuss in section 6. This paper finishes with some concluding remarks and avenues for further research.

2. Literature review

2.1. Form-focused instruction

FFI can be defined as "any pedagogical effort which is used to draw the learners' attention to language form either implicitly or explicitly [...] within meaning-based approaches to second language (L2) instruction [and] in which a focus on language is provided in either spontaneous or predetermined ways" (Spada, 1977, p. 73). It has often been characterised in terms of its explicitness (see Spada, 2023) and, although both implicit and explicit FFI can be beneficial for L2 learning, the latter has been observed to be more advantageous in meta-analyses (i.e. secondary research aimed at identifying cumulative results arising from a body of primary studies (Roehr-Brackin, 2018, p. 91)) of the effect of different types of L2 instruction such as Norris and Ortega (2001) and Goo et al. (2015), particularly in the short run, and in the development of explicit knowledge. This type of FFI comprises techniques such as L1-L2 contrasts and explanations, among others.

Research on FFI is currently immersed in exploring different types of FFI in different contexts (Nassaji, 2016) and in whether FFI is more effective or necessary for certain forms, particularly in the case of young learners, a population not receiving attention until quite recently (see Collins & Ruivivar, 2020, 2021; Roehr-Brackin, 2024, in this respect). The underlying reason for the scarcity of research along these lines with children may be related to the belief that children learn implicitly, which renders FFI

unsuitable. However, 10- to 12-year-olds show the capacity for abstract thought and are able to focus on linguistic issues (Hakes, 1980). Similarly, recent research suggests that metalinguistic abilities may increase at ages 8–9 when children are provided with age-appropriate FFI (Roehr-Brackin & Tellier, 2019), and that explicit instruction may affect children and adults in the same way (Ferman & Karni, 2010; Lichtman, 2021).

The ample majority of studies on the effect of FFI have targeted learners from Canadian immersion programmes who, despite the large amount of exposure received, lagged behind in terms of the accurate production of certain features, such as past tense/verb aspect (Harley, 1989), grammatical gender in French (Harley, 1998), L2 English question formation (White et al, 1991), and English possessive determiners (White & Ranta, 2002), among others. These investigations proved the effectiveness of various FFI treatments on the development of explicit knowledge of these linguistic features. Nevertheless, as suggested in current research (see Hanan, 2015), the effectiveness of these treatments could have been strengthened by the large exposure to meaningful input offered in immersion contexts.

Evidence from input-limited contexts is less abundant. One of those earlier studies in these contexts was Hanan (2015) with English-speaking learners of German. The combination of explicit information about the target features selected, together with structured input activities, led to gains in learners' ability to use the accusative case in masculine definite articles in German. However, this investigation did not include an explicit-only condition, and consequently, the extent to which explicit information about the target feature contributed to learning remained unknown.

Gorman and Ellis (2019) pioneered in this case by providing young English as a foreign language (EFL) learners (9–12-year-olds) with metalinguistic information in between performing communicative tasks. They also examined direct written corrective feedback as another FFI intervention not previously tested with young learners. In particular, they compared two experimental condition groups (ME and direct written corrective feedback) to a dictogloss+self-correction group who did not receive any treatment at all but were allowed only to self-correct. The study lasted for five weeks in which the three groups completed four dictogloss tasks individually targeting the present perfect simple, a novel feature to the participants. In week 1, the three groups performed the first dictogloss (as a pre-test). In week 2, the treatment received by the dictogloss+ME group consisted of an eight-minute mini-lesson one week after the performance of the first dictogloss task, while the second treatment group were provided eight minutes to study their direct written corrective feedback. The dictogloss+self-correction group were given eight minutes to edit their previous production, not receiving any type of feedback. Participants from the experimental and comparison groups were then asked to perform the second dictogloss. This

procedure was subsequently repeated in week 3 when the three groups performed the third dictogloss. After week 4, in which learners did not perform any task, the three groups completed the fourth dictogloss (as a delayed post-test).

The examination of the results indicated that neither FFI in the form of dictogloss+ME nor dictogloss+direct written corrective feedback had an impact on the learners' accurate use of the target structure over the performance of a dictogloss task coupled with self-correction. As regards ME, the authors acknowledge the lack of practice and the need for more engaging ME for children. With respect to direct written corrective feedback, they suggest that learners may not have been developmentally ready to pay attention to the corrections made, or that they were too focused on reconstructing the content in new pieces of writing that prevented them from drawing upon the explicit knowledge they had gained during the treatment phase. Thus, these authors made a call for further research along these lines.

Other studies conducted in EFL settings have also tested individual FFI types (Calzada, 2021; Serrano, 2011). The study by Calzada (2021) with L1 Spanish 6th grade primary school learners (11–12 years old) examined the effect of pre-task instruction on the accurate use of the third person singular *-s* morpheme and the possessive determiners *his/her* in a dictogloss task. To this end, one group received two 10-minute pre-task focus-on-form mini-lessons to raise learners' awareness of the third person singular *-s* morpheme and the possessive determiners, respectively. These mini-lessons were delivered following the same sequence of activities before the dictogloss task: input enhancement, elicitation and provision of the rule, and practice. Another group just performed the two dictogloss tasks collaboratively, while a third group carried out these tasks individually.

The examination of the reconstructed texts showed that the group with pre-task focus-on-form mini-lessons was more accurate in the use of the target forms, but the mean accuracy score still remained somewhat low across groups (below 50%). The analysis of the results in terms of LRE production also confirmed greater benefits of the pre-task focus-on-form mini-lessons in the case of the third person singular *-s* morpheme than in the case of the possessive determiners. As explained by the author, this finding could be ascribed to (1) the existence of more instances of the third person singular *-s* morpheme in the input text, which could have resulted in greater opportunities for languaging, (2) to the design of the task which could have prompted the omission of possessive determiners or the substitution by 'the', as well as (3) the lesser amount of instruction and low frequency of this feature in the input.

Other studies targeting the English possessive determiners (Serrano, 2011) have explored the effect of FFI in the form of metalinguistic explanations and practice.

Two groups of primary school learners immersed in a Content and Language Integrated Learning (CLIL) programme participated in the study. One group received metalinguistic instruction for 6 weeks, while a control group did not. The three groups completed three different tasks at a pre- and a post-task stage: a metalinguistic judgement task to look into whether learners demonstrated knowledge of the target features, as well as an error correction task and a production oral task to measure their performance. The results obtained indicated that the treatment received by the experimental group did not have an impact on metalinguistic knowledge as measured by the metalinguistic judgement task, but a slightly positive impact on the performance of the controlled task (error correction) and the production oral task. The positive effect of metalinguistic instruction was suggested to be dependent on individual differences, such as learners' analytic ability.

Even if FFI has attracted the attention of researchers targeting young learners in input-limited settings in recent years, investigations that comprise a wider range of FFI techniques, including additional FFI treatments to the performance of focused tasks, are needed. Some of the investigations already conducted did not yield significant improvements on the selected targets (i.e. Gorman & Ellis, 2019), in others subtle improvements were observed, and greater benefits were attested for certain features (e.g. third person singular *-s* morpheme) more than others (possessive determiners *his/her*) (i.e. Calzada, 2021).

This paper will try to contribute along these lines by examining two forms of FFI (dictogloss tasks+ME and dictogloss tasks+self-correction) following Gorman and Ellis (2019), though with the incorporation of three modifications which could reinforce the learning potential of a dictogloss task for the target feature selected (possessive determiners *his/her*): (1) the administration of a collaborative task instead of an individual dictogloss task, which has proved effective in a variety of contexts (Basterrechea & García Mayo, 2014; Calzada, 2021), and which could be more appropriately integrated in a meaning-oriented learning context, such as CLIL educational programmes, in which participants from the current investigation are enrolled; (2) the inclusion of an engaging treatment with a noticing and an awareness phase followed by practice along Bouffard and Sarkar (2008), and Lyster (2007, 2015); and (3) the incorporation of precise instructions for learners to draw their attention to form rather than meaning during task performance.

The examination of the use of possessive determiners in the written output will allow us to assess the differential effect of two forms of FFI on the application of the explicit knowledge in productive language use. As a written task, characterised by a more self-determined pace (Kuiken & Vedder, 2012), and as a task performed in collaboration, the assumption is that this task generates reflection about language

(Erlam, 2013; Gutiérrez, 2013; Thepseenu & Roehr, 2013), and thus the possibility of drawing upon explicit knowledge about possessive determiners facilitated by explicit instruction in the form of ME, all of which could lead to greater accuracy for these language forms in the ME group.

2.2. Language-analytic ability

Language-analytic ability, being a component of language learning aptitude, has been defined as the ability to infer linguistic systematicities from the input and make generalisations, encompassing both grammatical sensitivity and inductive language learning ability (Skehan, 1998, 2002). Previous studies have mostly explored its relationship with grammar development in adults, with a meta-analysis concluding that this ID is a good predictor of grammar learning over other components of aptitude (Li, 2016). Likewise, recent studies have also confirmed that this ID is a predictor of success for the acquisition of grammar properties in children (Roehr-Brackin & Tellier, 2019; Rosa-González, 2011).

Latest research has also been immersed in how FFI interacts with different IDs (DeKeyser, 2021; Lado & Sanz, 2021; Li, Ellis, & Zhu, 2019; Pawlak, 2021; Suzuki, 2022). For language-analytic ability in particular, a compensation pattern where FFI inhibits the effects of the ID, being equally effective for all students, has been observed in adults (Suzuki, 2019; Suzuki, Yokosawa, & Aline, 2022) and adolescents (White & Ranta, 2002). As Suzuki (2022) claims, "extra support (e.g., the aid of explicit information) can level out learners' differences" (p. 299). In this regard, current investigations are concerned with the interaction between this ID and different types and combinations of FFI (Kachinske & DeKeyser, 2019; Li et al., 2019) as it remains unknown if such a compensation pattern will be observed with any FFI implementation (see Serrano, 2011 for an example where a compensation pattern between FFI in the form of ME only and language-analytic ability was not found).

Since language-analytic ability has also been found to be a good predictor in children as young as 8 years old (Roehr-Brackin & Tellier, 2019; Rosa-González, 2011), examining the interaction between different types of FFI and this ID in young learners will undoubtedly contribute to uncover the optimal combination of individual and contextual factors needed to maximise the benefits of explicit learning and teaching in this population. Considering the scarcity of studies that have undertaken this research (e.g. Li et al., 2019) and the recent calls to compare different teaching approaches in the same age group, or to compare same-age children with different levels of language learning aptitude participating in the same teaching approach (Roehr-Brackin, 2022, 2024), the present investigation will try to fill these gaps by exploring the interaction of language-analytic ability and FFI in two groups of young learners: one receiving FFI

in the form of focused tasks (i.e. dictogloss) coupled with self-correction, and another one in the form of ME in addition to the performance of the same focused tasks.

3. Research questions

Despite the increasing attention FFI is receiving in the case of young learners in limited-input settings, research is still in its infancy and more investigations targeting different types and combinations of FFI interventions are needed. Likewise, even if FFI has been reported to blur differences ascribed to language-analytic ability (Suzuki, 2022), research with young learners is still limited and, in particular, how different ways of FFI may affect the relationship between this ID and language achievement remains an open question. Thus, this study, integrated in a bigger project where different FFI conditions have been explored in intact classes as a response to a call for more authentic classroom research (Loewen & Sato, 2018; Philp, Walter, & Basturkmen, 2010; The Douglas Fir Group, 2016), will try to contribute to the aforementioned gaps by addressing the following questions:

1. Does FFI in the form of metalinguistic explanations after the performance of a focused task lead to greater accuracy in the use of the target feature selected in reconstructed texts than self-correction after the performance of a focused task?

2. How do different forms of FFI (dictogloss+metalinguistic explanations; dictogloss+self-correction) mediate the relationship between language-analytic ability and use of the target feature selected?

4. Methodology

4.1. Participants

Thirty-three Basque/Spanish bilinguals (20 girls and 13 boys) from two intact classes in 5th grade from a semi-private school voluntarily participated in this study after obtaining parental consent. All of them came from Spanish-speaking families, but thanks to their enrolment in a trilingual programme, students achieve a good command in L2 Basque, so they can be considered Basque/Spanish bilinguals dominant in L1 Spanish (Cenoz, 2009). They were between 10 and 11 years old (M=10.2, SD=0.4) and, on average, had started learning English at the age of 3 (M=3.3, SD=1.8). All of them claimed to have a perfect command of Spanish, a Romance language.

Learners were taking part in a trilingual CLIL programme where Spanish, Basque and English were the languages of instruction. CLIL lessons were characterised by the

lack of FFI interventions in their lessons. As regards content subjects, they were taking mathematics in Spanish; physical education, arts and music in Basque; and sciences, robotics and religion in English. In addition to this, they also received lessons that focused on the language itself: 3 times a week for English, and 4 times a week for Spanish and Basque each. In total, they were exposed to approximately 363 hours of English a year during class time. Outside of the school, 30% of students in each class attended private English lessons for 2 hours on average (M=2.1, SD=1.1). The characteristics of these participants reflect the regular profile of learners during middle childhood (7–11 years old) attending a semi-private school in a bilingual community in Spain, usually enrolled in a CLIL programme, and some of them attending private English lessons. Thus, the outcomes obtained in the present study could be potentially applicable to typically intact classes. At the time of testing, they had an A1-A2 level of English according to the Flyers Test (Cambridge Assessment English, 2018). The administration of this test allowed us to ensure that we had two comparable groups in terms of general language proficiency, as confirmed by a Mann-Whitney U test (z=1.087, p=0.290, d= 0.385). In terms of the target feature knowledge, they were also comparable as evidenced by the lack of statistically significant differences in a grammaticality judgement task (GJT) administered at the outset of the study (z=0.674, p=0.509, d=0.234). This task included 16 items related to the target feature in both gender-matched and gender-mismatched contexts as well as 12 distractors (see section 4.2 for an explanation of gender-matched and gender-mismatched contexts in the use of English possessive determiners; see Iglesias-Diéguez & Martínez-Adrián, in press for a more detailed analysis of the GJT).

Each of the two classes were randomly assigned to one of two conditions: the dictogloss+self-correction group (n=15; 10 girls and 5 boys), who received explicit FFI in the form of collaborative dictogloss tasks coupled with self-correction, and the dictogloss+ME group (n=18; 10 girls and 8 boys), who were involved in awareness-raising sessions in addition to the collaborative dictogloss tasks. To carry out the dictogloss, participants in each group were paired up considering their proficiency level as measured by the Flyers test, resulting in 6 pairs and 1 triad in the dictogloss+self-correction group and 9 pairs in the dictogloss+ME group. Having proficiency-matched pairs with children fosters collaborative patterns of interaction and allows them to pool their resources and remain on task (Basterrechea & Gallardo-del-Puerto, 2020; García Mayo & Imaz Agirre, 2019).

4.2. Target structure

So as to investigate the effect of FFI in the form of ME, we selected the English third person possessive determiners *his/her* as the target structure. Unlike Gorman and Ellis (2019), this feature is not novel to the participants from the present study,

but it has been selected on the grounds that it poses considerable difficulty to students with Romance L1s (White et al., 2007) due to the differences in gender assignment between Romance languages and English. Whereas the English possessive determiner depends on the gender of the possessor, Romance languages assign the gender of the possessive determiners based on the gender of the possessed entity. In particular, possessive determiners in Spanish do not bear any marking of gender, the determiners being equal regardless of the gender of the possessor or of the possessed entity (e.g. *su coche* 'his/her car'). However, number agreement occurs with the possessed entity (e.g. *su coche* 'his/her car', *sus coches* 'his/her cars'), and the gender assignment typical of Romance languages can still be seen in third person possessive pronouns (e.g. *el suyo* 'his/hers' [masc.], *la suya* 'his/hers' [fem.]). This difference in gender-agreement systems gives rise to two different contexts from the perspective of a learner of English with a Romance L1 background: (1) *gender-matched contexts*, when the gender of the possessive determiner agrees with the gender of the entity being possessed (e.g. *his grandfather*, *her aunt*); and (2) *gender-mismatched contexts*, when the gender of the possessive determiner differs from the gender of the entity being possessed (e.g. *his grandmother*, *her uncle*).

Previous studies emphasise the importance of the context as they report different accuracy scores by Basque/Spanish bilinguals in the production of English possessive determiners. While both contexts pose difficulty for learners of English, gender-mismatched contexts seem to be more challenging for speakers with a Romance L1 background (Imaz Agirre & García Mayo, 2013). Therefore, Basque/Spanish bilinguals need to overcome this crosslinguistic influence, especially in gender-mismatched contexts, to achieve target-like use of the possessive determiners in English. In this regard, note that Basque does not have grammatical gender. Therefore, the difficulties in dealing with the possessive determiners *his/her* in the case of Basque/Spanish bilinguals are ascribed to the influence of gender-assignment in Spanish, as has been documented in previous studies with L1/L2 Basque learners of English (Imaz Agirre & García Mayo, 2013).

4.3. Design

Two groups participated in four sessions as indicated in Table 1. In the first session, all participants completed part II of the MLAT-ES test (Stansfield & Reed, 2005) to measure their language-analytic ability. In session 2, both groups completed the first dictogloss task, which served as the pre-test to check their accuracy in the use of the possessive determiners. Session 3 was devoted to the round of treatment based on dictogloss 1 for the dictogloss+ME group (15 minutes) followed by the second dictogloss in both groups. To counteract any effect that the additional time with the outcome of dictogloss 1 may have in the dictogloss+ME group, pairs in the dictogloss+self-correction group were given 15 minutes to go through their previous

writing from dictogloss 1 and self-correct it. The same procedure was followed in session 4 with the round of treatment for the dictogloss+ME group based on dictogloss 2, the self-correction time for the dictogloss+self-correction group, and the completion of the third dictogloss in both groups, which was used as the post-test.

Table 1: Design of the study

	Dictogloss+self-correction group	**Dictogloss+ME group**
Session 1	MLAT-ES (Part II) (30')	MLAT-ES (Part II) (30')
Session 2	Dictogloss 1 [PRE-TEST] (10')	Dictogloss 1 [PRE-TEST] (10')
Session 3	Self-correction of dict. 1 (15') Dictogloss 2 (10')	Round 1 of ME treatment (15') Dictogloss 2 (10')
Session 4	Self-correction of dict. 2 (15') Dictogloss 3 [POST-TEST] (10')	Round 2 of ME treatment (15') Dictogloss 3 [POST-TEST] (10')

The design of this study builds on Gorman and Ellis (2019), although there are three important differences. First, the dictogloss task was carried out in pairs to foster students' focus on form as previous studies show that carrying out a written task collaboratively can be beneficial for language development (Fernández Dobao, 2012; Kim, 2008; Storch, 1999) and results in a positive experience for students (Fernández Dobao, 2020; Fernández Dobao & Blum, 2013). Secondly, as suggested by Gorman and Ellis, we attempted to make the treatment (i.e. ME sessions) more engaging by using students' own written production during these sessions (as in Bouffard & Sarkar, 2008) and varying the delivery method for the practice phase based on Lyster (2007, 2015). Thirdly, all participants were explicitly encouraged to produce grammatically accurate texts as part of the instructions for the dictogloss task, in order to avoid situations reported in Gorman and Ellis (2019), where learners were so focused on reconstructing content that it prevented them from drawing upon the explicit knowledge gained during treatment.

4.4. Instruments and materials

4.4.1. Dictogloss

The collaborative task that students had to carry out in pairs consisted in a dictogloss task (Wajnryb, 1990), where learners had to replicate (i.e. recall and reconstruct) a text in written form that they had previously listened to. This task has been characterised as a focused task as it is designed with the intention of drawing learners' attention to specific language features in a meaning-based context (Benati,

2021; García Mayo, 2002) and has been previously used when targeting the possessive determiners *his/her* (Calzada, 2021). For this study, three different texts were developed and recorded by an L1 English speaker. The texts provided familiar situations to our participants as they narrated a family's routines in a typical day in the garden, a day that they spend together on holidays and a day deciding and buying what they want after having won the lottery. The second dictogloss can be found in Appendix 1 as an illustration.

Considering the limitations acknowledged in previous studies examining production data and the acquisition of possessive determiners (Calzada, 2021), as well as the instructional context in which the study was conducted, these texts were kept as short as possible (107 words each) to avoid learners' attention focusing excessively on meaning rather than on form. Moreover, the number of instances of the target structure appearing in each text has been increased with respect to previous studies: each dictogloss contained 8 instances of possessive determiners, half of which were gender-matched and the other half gender-mismatched. These were obligatory contexts (OCs), where the possessive determiners could not be substituted or omitted. Moreover, the possessives in each of those contexts were split into an equal number of *his* and *her*. This resulted in a total of 2 gender-matched *his*, 2 gender-matched *her*, 2 gender-mismatched *his* and 2 gender-mismatched *her* in each text. Additionally, the texts were syntactically simple, making minimal use of embedded clauses and favouring coordination, and lexical items contained in the texts were appropriate for the level of our participants.

Each dictogloss was accompanied by a worksheet for each student where they could take notes. Unlike in previous studies, these worksheets contained 4 keywords that appeared in the texts to avoid omissions of the target structure or substitution by the definite article *the*, which has been observed in previous research when an explanatory image is included (Calzada, 2021). These keywords also contributed to reducing participants' cognitive load when trying to remember the content of the story.

4.4.2. MLAT

To measure participants' language-analytic ability, part II of the MLAT-ES in pen-and-paper format was used (Stansfield & Reed, 2005). The MLAT-ES is a modified version of the MLAT-E (Carroll & Sapon, 2002) for children specifically geared towards L1 Spanish young learners between the ages of 8 and 13. This test has proved reliable with children in a variety of settings (Kiss & Nikolov, 2005; Roehr-Brackin & Tellier, 2019; Tellier & Roehr-Brackin, 2013), including L1 Spanish children (Gesa & Suárez, 2022; Muñoz, 2014). Part II of the test intends to measure students' sensibility to grammatical structures avoiding explicit reference to grammatical terminology. To

do so, participants had to recognise the function of a given word in a sentence in Spanish and identify which word in a second sentence had the same function. This part of the MLAT-E has been used to measure language-analytic ability in previous studies (Roehr-Brackin & Tellier, 2019).

4.5. Procedure

Prior to the beginning of the study, all participants completed a background questionnaire that gathered data about the languages they knew and their participation in any extracurricular activities in English, such as private lessons.

Before starting with the sequence of dictogloss tasks, all participants completed the MLAT-ES. It contained 4 training items that were used to make sure all students had understood the instructions correctly. After this, they were given 30 minutes to complete the test.

In each dictogloss task, the dictogloss worksheet described above was handed out to each student. Before listening to the text, they were asked to read aloud the four keywords presented at the top of the page. If a participant did not understand the meaning of a keyword, they were told the meaning of the word in Spanish to facilitate comprehension of the text. After this, they listened to the text twice. The first time, they simply listened to it, whilst the second time they were allowed to take notes individually in their worksheets, although they were encouraged to avoid writing full sentences. Then, each pair was given 10 minutes to reconstruct the story using the keywords and sharing their notes. Both measures could reduce learners' cognitive load, facilitating the recall of the content (Leeming, Aubrey, & Lambert, 2022), and thus helping learners focus on formal aspects of the language during the reconstruction phase. They were also told that they had to produce a grammatically correct text.

The ME treatment consisted in 15-minute mini-lessons that took place in class immediately before starting dictogloss 2 and 3, and were delivered by one member of the research team. Based on Bouffard and Sarkar (2008) and Lyster (2007, 2015), each mini-lesson was structured into the same three phases: noticing, awareness-raising and practice. In the noticing phase, students were shown some sentences from the previous dictogloss that contained, at least, an error related to the possessive determiners. The purpose of including participants' own production was to foster their engagement in the treatment. The researcher asked the whole class if they could identify any mistakes in those sentences and elicited answers from different students. They were also asked to provide a correction for those mistakes.

After having gone through the examples, the awareness-raising phase began by posing questions to students about how possessive determiners work in English (e.g., *why do we say 'his' and not 'her' in some of the sentences we have corrected?*) as well as about the reason why they may be difficult for Spanish speakers (e.g., *why do you think we make these mistakes?*). During this phase, they were guided by the researcher towards a general rule to explain gender agreement in English, which was explicitly stated and shown to students before continuing to the practice phase.

This last phase consisted in a cloze activity, where students had to select the correct possessive determiner to complete several sentences. In the first mini-lesson (before dictogloss 2), students had to raise a flashcard with the correct possessive determiner to complete the sentences that they were shown in the whiteboard, whereas in the second mini-lesson (before dictogloss 3), the activity was carried out through *Kahoot!*, an audience response tool, using students' tablets. We varied the way of delivering the practice in an attempt to make the activity more fun and dynamic for young learners. Except for the practice phase in the ME mini-lessons, the dictogloss+ME group did not receive any feedback on their production or any feedback related to the possessive determiners.

In the dictogloss+self-correction group, participants did not receive any specific instruction on the possessive determiners. Likewise, this group did not receive any feedback on their dictogloss production or on their self-correction of the text, as in Gorman and Ellis (2019). To compensate for the extra time that the dictogloss+ME group spent engaging with their production from the previous dictogloss, the dictogloss+self-correction group was also given 15 minutes to self-correct their previous writing. The self-correction session also took place immediately before starting a new dictogloss task to maintain comparability between the groups.

4.6. Data coding and analysis

Participants' change in the accuracy of the use of the possessive determiners was obtained in the following way. First, the OCs for the target structure were identified and counted in students' texts. When the correct determiner was provided in an OC, it was counted as an instance of correct use of the target structure. If a possessive determiner was incorrectly provided in a context where it was not needed, it was coded as an instance of overproduction. With this information, Pica's (1983) formula was used to obtain a percentage for the target-like use (TLU) of the possessive determiner for each participant in each dictogloss:

$$\text{“TLU”} = \frac{\text{“instances of correct use in an OC”}}{(\text{“number of OCs”} + \text{“instances of overproduction”})}$$

The TLU is considered a valid measure of accuracy in both written and oral production (Ellis & Barkhuizen, 2005) and has been used in previous studies with this purpose (e.g. Akbaş & Ölçü-Dinçer, 2021; Frear & Chiu, 2015; Vraciu, 2020).[2]

Finally, the difference between the TLU in dictogloss 3 and the TLU in dictogloss 1 for each participant was obtained as a measure of the development in students' accuracy in the use of the target structure. To control for the influence of prior knowledge students may have, gains were used instead of raw scores when comparing the dictogloss+self-correction and dictogloss+ME groups (File & Adams, 2010; Xu & Li, 2021; Zimmerman & Williams, 1982). Although we acknowledge that participants with higher accuracy in the pre-test can obtain smaller gains, there were no statistically significant differences in the distribution of high and low scores in the pre-test (z=-0.606, p=0.62, d=0.311) between the two groups.

As regards the MLAT-ES, a point was awarded for each correct answer. The maximum score that could be obtained in part II of this test was a total of 30 points. There was no difference in the distribution of MLAT scores between the two groups (z=0.911, p=0.381, d=0.319).

Although the Kolmogorov-Smirnov tests for normality were not significant for some variables in each group, exploration of the histograms and Q-Q plots revealed deviance from normality in the residuals. Therefore, non-parametric tests were used in all instances. Cohen's *d* effect size values were calculated, and subsequently interpreted following the specific benchmarks for L2 acquisition proposed by Plonsky and Oswald (2014) for both intergroup comparisons (values around 0.4 were considered small, medium around 0.7 and large if around 1.0) and intragroup comparisons (values around 0.6 were considered small, medium around 1.0 and large if around 1.4).

To answer the first research question, which looked into the effect of ME on the accurate use of *his/her* in gender-matched and gender-mismatched contexts, we analysed the difference in the accuracy of the possessive determiners in dictogloss 1 (i.e. pre-test) and dictogloss 3 (i.e. post-test) in each group, as well as the difference in gains between the groups. To do so, we calculated the mean TLU for each group using the TLU for each participant and implemented a Wilcoxon test or a Mann-Whitney U test to reveal significant differences in global accuracy in the use of the possessive determiners.[3] Results were also split by the context (i.e. gender-matched and gender-mismatched) but only descriptive statistics are reported due to the low number of instances produced in some cases, which prevents the use of inferential analyses.

To answer the second research question, which focused on the interplay between language-analytic ability and FFI, Spearman's rho was calculated in each group to

test the association between students' language-analytic ability and the gains they obtained in global accuracy measured via their differences in TLU between dictogloss 1 and 3. Associations between participants' language-analytic ability and their gains in both gender-matched and gender-mismatched contexts were analysed by visual inspection of the scatterplots and identification of trends as the number of data points was inadequate for inferential statistics.

5. Results

The first analysis aimed at answering the first research question by exploring the effectiveness of ME combined with collaborative dictogloss tasks on learners' accuracy in the use of the possessive determiners *his/her* as compared to the combination of the same tasks with self-correction (see Tables 2 to 4). Note that accuracy can only be measured in those pairs that produced OCs, and that the TLU was calculated for each pair that produced OCs and then averaged out to obtain a mean TLU for each group in dictogloss 1 and in dictogloss 3, rather than aggregating the data from all pairs yielding one single TLU for each group with no SD. Doing so allows for a more accurate representation of each pair's performance and their variability within each group.

Table 2: Global descriptives of obligatory contexts (OCs), correct provisions of the target feature, overproduction and target-like use (TLU) for dictogloss 1 and 3 in both groups

		Dictogloss+self-correction group (n=7 pairs)		Dictogloss+ME group (n=9 pairs)	
		Dictogloss 1	Dictogloss 3	Dictogloss 1	Dictogloss 3
OCs (N)		21 (7 pairs)	23 (7 pairs)	26 (7 pairs)	24 (6 pairs)
Correctly produced (N)		13	10	14	18
Overproduction (N)		0	0	0	0
TLU (%)	Mean	70.2	44.0	64.3	90.6
	SD	22.5	14.2	26.2	10.4
	Min	50	25.0	33.3	80
	Max	100	66.7	100	100
	Range	50	41.7	66.7	20

Table 3: Descriptives of obligatory contexts (OCs), correct provisions of the target feature, overproduction and target-like use (TLU) in gender-matched contexts for dictogloss 1 and 3 in both groups

		Dictogloss+self-correction group (n=7 pairs)		Dictogloss+ME group (n=9 pairs)	
		Dictogloss 1	Dictogloss 3	Dictogloss 1	Dictogloss 3
OCs (N)		11 (5 pairs)	18 (7 pairs)	10 (5 pairs)	10 (5 pairs)
Correctly produced (N)		8	7	9	8
Overproduction (N)		0	0	0	0
TLU (%)	Mean	90.0	36.9	80.0	88.3
	SD	22.4	34.6	44.7	16.2
	Min	50	0	0	66.7
	Max	100	100	100	100
	Range	50	100	100	33.3

Table 4: Descriptives of obligatory contexts (OCs), correct provisions of the target feature, overproduction and target-like use (TLU) in gender-mismatched contexts for dictogloss 1 and 3 in both groups

		Dictogloss+self-correction group (n=7 pairs)		Dictogloss+ME group (n=9 pairs)	
		Dictogloss 1	Dictogloss 3	Dictogloss 1	Dictogloss 3
OCs (N)		10 (6 pairs)	5 (4 pairs)	16 (6 pairs)	14 (6 pairs)
Correctly produced (N)		5	3	9	13
Overproduction (N)		0	0	0	0
TLU (%)	Mean	50.0	75.0	52.8	95.8
	SD	45.9	50.0	32.3	10.2
	Min	0	0	0	75
	Max	100	100	100	100
	Range	100	100	100	25

In the dictogloss+self-correction group, all 7 pairs produced OCs in dictogloss 1 (M=3.0, SD=1.83) and in dictogloss 3 (M=3.29, SD=1.38). As regards their accuracy in the use of the possessive determiners, displayed in Table 2, students achieved a TLU of 70.2% in dictogloss 1 (SD=0.225), which fell down to 44.0% in dictogloss 3 (SD=0.142). A Wilcoxon's signed-rank test shows that the scores in post-test are significantly lower (z=-2.032, p=0.042), with a large effect size (d=1.294).

When the context of the possessive determiners is considered, some differences emerge between gender-matched and gender-mismatched contexts in both the production of OCs and the accuracy of use. In gender-matched contexts (Table 3), 5 pairs produced OCs (M=2.20, SD=2.17) with an average accuracy of 90.0% (SD=0.224) in dictogloss 1, whereas all 7 pairs produced OCs in dictogloss 3 (M=2.57, SD=1.9), with only a 36.9% accuracy (SD=0.346). In gender-mismatched contexts (Table 4), 6 pairs produced OCs in dictogloss 1 (M=1.67, SD=1.03) with an average accuracy of 50% (SD=0.459), whereas only 4 pairs produced OCs in dictogloss 3 (M=1.25, SD=0.5), with a TLU of 75% (SD=0.5). More pairs produced OCs in dictogloss 3 than in dictogloss 1 in matched contexts, but the opposite trend is observed in mismatched contexts. Moreover, accuracy decreased in gender-matched contexts and increased in gender-mismatched contexts.

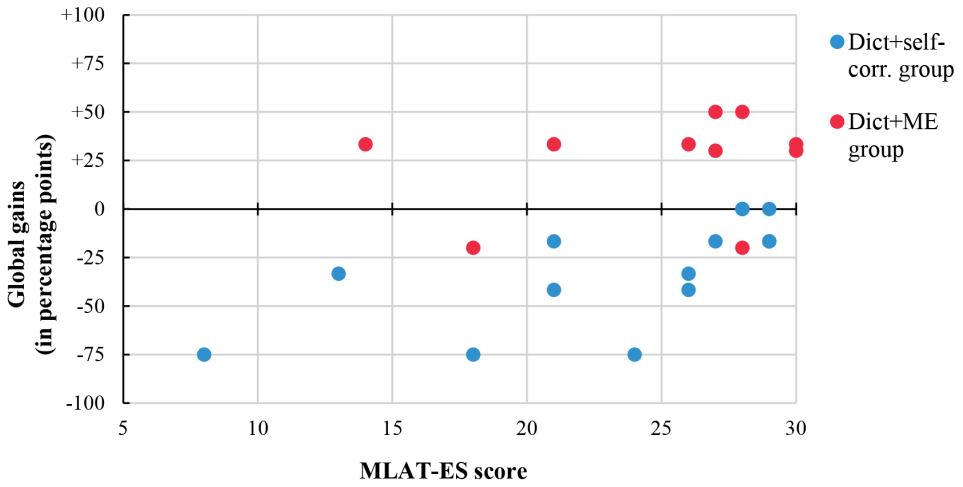
In the dictogloss+ME group, 7 out of 9 pairs produced OCs in dictogloss 1 (M=3.71, SD=2.56) and 6 pairs in dictogloss 3 (M=4.0, SD=1.55). As regards their accuracy in the use of the possessive determiners, displayed in Table 2, students achieved an average TLU of 64.3% in dictogloss 1 (SD=0.262). Although a Wilcoxon's signed-rank test shows that the post-test scores were not significantly higher than the pre-test scores (z=-1.761, p=0.078), accuracy rose to 90.6% in dictogloss 3 (SD=0.104), a result that could be reinforced by the existence of a large effect size (d=1.341).

Some differences can also be observed in the dictogloss+ME group when the context of the possessive determiners is considered. In gender-matched contexts (Table 3), 5 out of the 7 pairs produced OCs (M=2.0, SD=2.24) with an average accuracy of 80.0% (SD=0.447), and 5 pairs produced OCs in dictogloss 3 (M=2.0, SD=1.41), with an 88.3% accuracy (SD=0.162). In gender-mismatched contexts (Table 4), 6 pairs produced OCs in dictogloss 1 (M=2.67, SD=0.82) with an average accuracy of 52.8% (SD=0.323), and 6 pairs produced OCs in dictogloss 3 (M=2.33, SD=1.03), with a TLU of 95.8% (SD=0.102). Contrary to the dictogloss+self-correction group, there was no increase or decrease in the number of pairs that produced OCs in any of the contexts in the dictogloss+ME group. This was accompanied by a slight improvement in accuracy in gender-matched contexts and a more prominent increase in accuracy in mismatched contexts, achieving highly accurate production of the possessive determiners in both contexts.

As regards the impact of the metalinguistic explanations on the gains in the use of the possessive determiners *his/her*, a Mann-Whitney U test on the gains for the pairs that had produced OCs in *both* dictogloss shows that the gains in the dictogloss+ME group are significantly greater than the gains in the dictogloss+self-correction group (z=2.204, p=0.03, d=1.635), with a large effect size. In particular, the pairs in the dictogloss+self-correction group that produced OCs in both dictogloss tasks had an average gain of -26.2 percentage points (SD=0.265), whereas the pairs that received the ME treatment gained on average +25.3 percentage points (SD=0.265).

When we break it down considering the context of the possessive determiners, we can see that the average gains in the dictogloss+self-correction group were -53.3 (SD=0.361) percentage points in matched contexts and +11.1 (SD=0.839) in mismatched contexts. In the dictogloss+ME group, the average gains were +22.2 (SD=0.385) in matched contexts and +31.7 (SD=0.325) in mismatched contexts. In the dictogloss+self-correction group, negative gains are observed in matched contexts and positive gains in mismatched contexts, although these are lower than the dictogloss+ME group, where gains are always positive as well as more stable across contexts.

Graph 1: Scatterplot with the individual scores of the MLAT-ES in the horizontal axis, and the gains from dictogloss 1 to dictogloss 3 (irrespective of the context) in the vertical axis

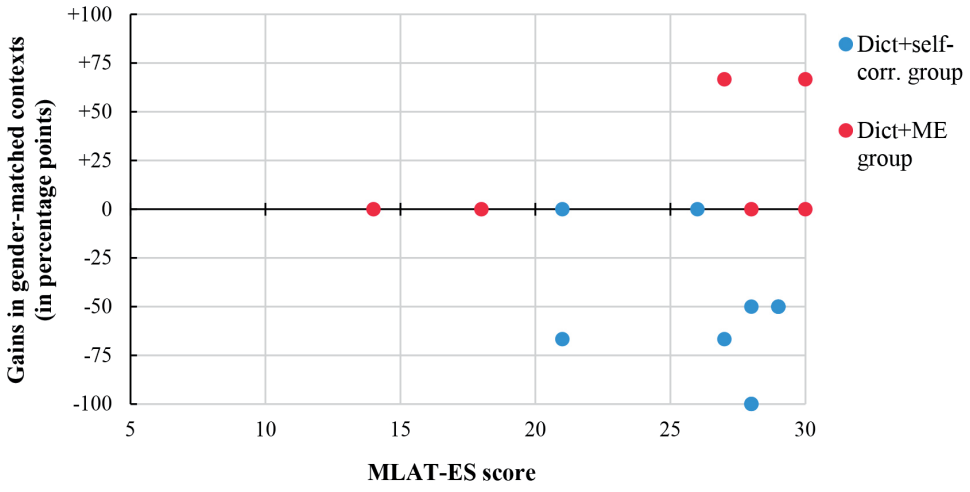


The second research question probed into the role of language-analytic ability in participants' gains in accuracy (or lack thereof). The Spearman's rho test reveals a strong and positive association between language-analytic ability and the gains

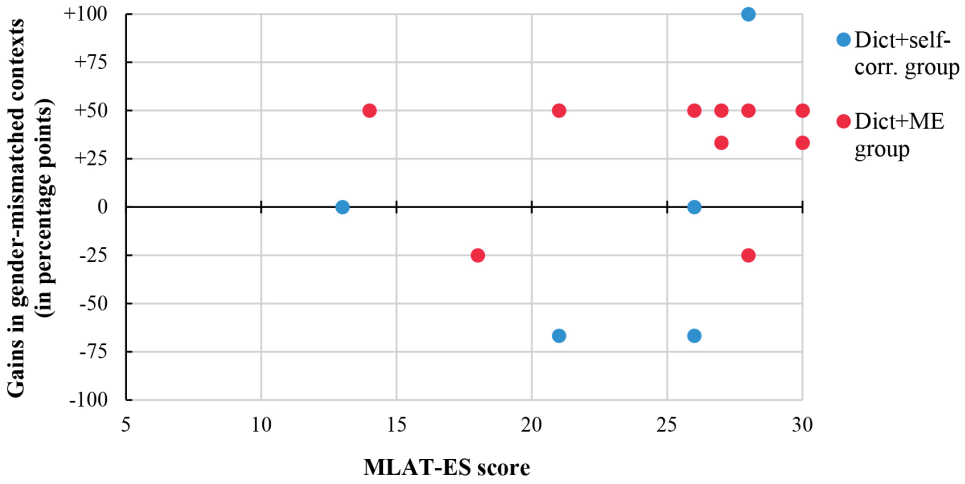
obtained in the dictogloss+self-correction group ($r_s(13)=0.755$, $p=0.001$), whereas no such association is found in the dictogloss+ME group ($r_s(8)=0.019$, $p=0.958$). As can be seen in Graph 1, participants in the dictogloss+self-correction group had smaller losses when their language-analytic ability was greater. In contrast, participants in the dictogloss+ME group generally obtained positive gains regardless of their language-analytic ability.

These trends seem to be driven by the influence of language-analytic ability on students' awareness of mismatched contexts. As shown in Graphs 2 and 3, we can observe that language-analytic ability seems to have a lesser impact in gender-matched contexts as gains are stable in both groups, although mostly negative in the dictogloss+self-correction group and mostly positive in the experimental group. In the mismatched context, however, gains remain stable and positive in the dictogloss+ME group, but they seem to follow a slightly positive association with their language-analytic ability in the dictogloss+self-correction group, although data points in this context are scattered and scarce. In the dictogloss+self-correction group, language-analytic ability seems to have some influence on accuracy in gender-mismatched contexts, or at least in their increased awareness of possessive determiners in this context due to the task.

Graph 2: Scatterplot with the individual scores of the MLAT-ES in the horizontal axis, and the gains from dictogloss 1 to dictogloss 3 in matched contexts in the vertical axis



Graph 3: Scatterplot with the individual scores of the MLAT-ES in the horizontal axis, and the gains from dictogloss 1 to dictogloss 3 in gender-mismatched contexts in the vertical axis



6. Discussion

In this section, results will be discussed according to the two research questions addressed. With respect to the first research question (*Does FFI in the form of metalinguistic explanations after the performance of a focused task lead to greater accuracy in the use of the target feature selected in reconstructed texts than self-correction after the performance of a focused task?*), the selected dictogloss task has been effective in drawing learners' attention to the possessive determiners in matched and mismatched contexts, the target features selected that have been found to pose difficulty for students with Romance L1s (White et al., 2007). Dictogloss tasks are focused tasks, characterised as explicit (Benati, 2021), which have proved efficient in drawing learners' attention to formal aspects of the language in a variety of contexts, with both adults and children (Alegría de la Colina & García Mayo, 2007; Basterrechea & García Mayo, 2014; Calzada, 2021; Lim & Jacobs, 2001).

Its efficiency, though, has been reported to depend on the type of target selected, yielding better results in the case of features such as the third persons singular *-s* morpheme than in the possessive determiners *his/her* (see Calzada, 2021).[4] In this regard, the fact that a greater number of targets for the different conditions were included in the three dictogloss tasks administered, the use of keywords rather than pictures to help the learners in their understanding of the text and to foster greater

attention to the language instead of the content, and the control for text length and vocabulary, allowing for more focused discussions, may have contributed to increase learners' attention to these features and to promote the production of OCs. In other words, the modifications introduced to solve the limitations observed in prior research, that were pilot-tested before the study began, have made the dictogloss an effective instrument for the elicitation of the possessive determiners *his/her*.[5]

The dictogloss task has been specifically effective in gender-mismatched conditions in both groups, which entail greater difficulty to learners with Romance L1s (White et al., 2007). In the case of the dictogloss+self-correction group, just performing the task without having received metalinguistic explanations can raise these learners' awareness of gender-mismatched contexts (see also Li, Ellis, & Kim, 2018), generating greater accuracy in this condition than in gender-matched contexts, where losses were observed. In this last respect, a restructuring process probably induced by the hypothesis-testing of language forms during the reconstruction phase may have taken place (Izumi, 2003; Swain, 1997; Swain & Lapkin, 1995), destabilising the possessive determiners in gender-matched contexts that generally seem to entail less difficulty for L1 Spanish speakers. In contrast, ME increases learners' awareness in both contexts, leading to an upsurge in accuracy in both gender-mismatched and gender-matched conditions and to an integration of both language forms. The mini-lessons of noticing, awareness and practice provided to our participants revolved around gender-matched and gender-mismatched contexts, and occurred after their production, preparing them for the next output-producing session. In other words, learners could have resorted to the knowledge gained during the ME sessions in their subsequent metalinguistic reflection (Swain, 1995) while reconstructing the texts, contributing to greater accuracy in the use of possessive determiners during production (Johns, 2003; Mitchell, 2000; Thepseenu & Roehr, 2013) than the participants in the self-correction group.

In this case, gains for the ME group could be considered robust enough in the light of the large effect size found when being compared to those of the self-correction group. Thus, unlike Gorman and Ellis (2019), explicit FFI has been found to be effective in the case of young learners for a feature not novel to them but that entailed learning difficulty due to the interaction between crosslinguistic influence and the developmental route for possessive determiners. We tentatively argue that adding a more engaging treatment with a noticing and an awareness phase followed by practice, giving precise instructions to focus on form, and asking them to perform the task collaboratively could foster the effectiveness of this combination of explicit forms of FFI.

Likewise, in contrast to other investigations in which explicit FFI has been found to be effective, low accuracy scores notwithstanding (below 50%) (e.g., Calzada, 2021; Serrano, 2011), in the present study, accuracy rose to 90% on average in the third

dictogloss task in the dictogloss+ME group. Thus, the addition of ME in the form of noticing and awareness sessions followed by practice proves as a better option to maximise the learning potential of these focused tasks than those employed in prior studies, a result reinforced by the examination of language development via a grammaticality judgement task administered to the same participants two weeks after the intervention. In particular, the group receiving ME has been found to show an average improvement of 12 points in the grammaticality judgement task when the global gains from the pre-test to the post-test were considered (Dictogloss+self-correction = +1.2; Dictogloss+ME = +12.3), as well as when examining gender-matched (Dictogloss+self-correction = +2.3; Dictogloss+ME = +11.3) and gender-mismatched contexts (Dictogloss+self-correction = +0.0; Dictogloss+ME = +13.2) separately (see Iglesias-Diéguez & Martínez-Adrián, in press). Further analyses await to determine whether the explicit knowledge gained thanks to the treatment received will last over time, considering the short duration of the treatment and the fact that explicit knowledge may be acquired faster but atrophies easily (Shintani & Eltis, 2013). This line of research will also respond to the call made in the literature by different authors (Calzada & García Mayo, 2020; Roehr-Brackin, 2018) for more studies on the durability of the benefits of FFI treatments that incorporate delayed post-tests. In this regard, recent studies conducted with children have unveiled benefits of FFI treatments in delayed post-tests (Li et al., 2018).

With respect to the second research question (*How do different forms of FFI (dictogloss+metalinguistic explanations; dictogloss+self-correction) mediate the relationship between language-analytic ability and use of the target feature selected?*), in the absence of a more guided treatment (i.e. dictogloss+self-correction group), this ID was found to be relevant, those students with a higher language-analytic ability attaining smaller losses (Roehr-Brackin & Tellier, 2019). In contrast, the more explicit the treatment received, the less relevant this ID was. In other words, learners with lower language-analytic ability may compensate their deficits thanks to the ME treatment received, a result in line with prior research conducted with both adults (Suzuki, 2019; Suzuki et al., 2022) and adolescents (J. White & Ranta, 2002). Further studies await to see whether this compensation pattern could be ascribed to an increase in the learners' language-analytic ability thanks to the treatment received, taking into account that this ID is not stable and could develop under FFI conditions (Roehr-Brackin & Tellier, 2019). Despite so, the present investigation has contributed to aptitude-treatment interaction research, a line of study that has started to thrive in the last years (DeKeyser, 2019; Suzuki, 2022). These results also seem to align with prior research in which this ID was more relevant with implicit types of FFI and less relevant with more explicit treatments (Li, 2013), all of which makes a step further as to "when and how L2 grammar instruction should be provided to learners with different strengths and weaknesses profiles of their cognitive abilities" (Suzuki, 2022, p. 297).

7. Conclusion

This paper has contributed to shed more light on the implementation of FFI in young learners in input-limited settings. More specifically, it has shown how FFI consisting of ME after the performance of a dictogloss task including a noticing and an awareness phase followed by practice along Bouffard and Sarkar (2008), and Lyster (2007, 2015) has led to greater awareness and increased accuracy in the use of the possessive determiners *his/her*. Likewise, results seem to indicate that the benefits of the FFI treatment received are independent of participants' language-analytic ability, blurring individual differences in this respect.[6]

For further research, it would be convenient to gather data from a bigger sample and to conduct the study in other learning contexts so as to increase the generalisability of the results obtained. Likewise, even if the combination of dictogloss tasks+ME has been effective in the promotion of accuracy, measures that could lead to a higher production of OCs in the case of possessive determiners would be worth exploring. Unlike the third person singular *-s* morpheme, features such as articles and possessive determiners have been found to be more difficult to elicit via focused tasks, as reported in previous research (Calzada & García Mayo, 2020, 2023; Collins & White, 2014). Adding a pre-task stage on family and friends' relationships could boost the task potential to elicit a greater number of forms and, in turn, to strengthen the value of the production data obtained. Similarly, considering that language-analytic ability is an *individual* difference, future studies including collaborative tasks could control for this variable by matching dyad members in terms of the scores obtained in the MLAT-ES test.[7] However, it should also be borne in mind that some studies have suggested that, regardless of the matching criterion, collaborative tasks can still be beneficial to young learners in a number of ways (Oliver & Bogachenko, 2019; see Oliver, Philp & Duchesne, 2017 for an example) and, thus, students need not be paired up by every single individual difference under study. All in all, since previous research has already unveiled the benefits of having proficiency-matched pairs (Basterrechea & Gallardo-del-Puerto, 2020; García Mayo & Imaz Agirre, 2019), future investigations, like the present study with intact classes, could focus on the effects of individual differences in proficiency-matched pairs as well, possibly in a classroom setting. This will facilitate the implementation of the pedagogical implications by practitioners as such pairings are arguably easier to carry out in a real classroom with the usual tests that students have to complete in class and, therefore, more ecologically valid.

Additionally, the incorporation of tasks tapping explicit and implicit knowledge would result in a more robust investigation of FFI effects in young learners. Qualitative measures such as retrospective interviews that could complement the examination

of LRE production also deserve attention as they could shed more light on covert metalinguistic activity (Gutiérrez, 2013).[8] Finally, longitudinal studies encompassing a wider array of FFI techniques and linguistic features (i.e. novel or already known but posing learning difficulty to the learners) are also desirable.

Despite the existing limitations, two main pedagogical implications are worth mentioning and discussing. First, primary school teachers might be ready to adopt focused tasks such as dictogloss if they want to draw learners' attention to linguistic features that pose greater difficulty for the English learner as for example the possessive determiners *his/her*. We want to stress the potential of these tasks that have already proved effective in adults and adolescent learners (Storch, 2016; Zhang & Plonsky, 2020), and are gaining momentum in the literature on child learners (Ellis, Li, & Zhu, 2019; Calzada & García Mayo, 2020, 2023). In addition, these tasks have been shown to be appealing to young EFL learners (Kopinska & Azkarai, 2020), all of which upholds their feasibility and suitability in this learning context. Furthermore, adding ME in the form of noticing and awareness sessions after the performance of dictogloss tasks in which real examples from the learners' production are shown can maximise the learning potential of these tasks, as they can enhance and induce the attention to linguistic features seeded in the text and enrich learners' metalinguistic reflection during text-reconstruction. In this vein, this is congruent with recent research that has confirmed the efficacy of explicit instruction within task conditions (Michaud & Ammar, 2023; Quinn, 2014), as learners can use the information provided during the treatment sessions in the follow-up tasks. Second, the addition of ME treatments is even more important in this learning context if we consider their potential to benefit students equally, regardless of their language-analytic ability. Thus, not only can the incorporation of this type of treatments promote learning opportunities, but also make them a reality to a greater number of students.

Acknowledgements

We would like to thank Andrea Robin Ruthven for recording the audio for the texts employed in the dictogloss tasks. This work was supported by the Spanish Ministry of Science and Innovation (PID2020-113630-GB-I00), the Basque Government (IT1426-22) and the University of the Basque Country (PIF20/300).

8. References

Akbaş, E., & Ölçü-Dinçer, Z. (2021). Accuracy order in L2 grammatical morphemes: Corpus evidence from different proficiency levels of Turkish learners of English. *Studies in Second Language Learning and Teaching*, *11*(4), 607-627. https://doi.org/10.14746/ssllt.2021.11.4.6

Alegría de la Colina, A., & García Mayo, M. P. (2007). Attention to form across collaborative tasks by low-proficiency learners in an EFL setting. In M. P. García Mayo (Ed.), *Investigating tasks in formal language learning* (pp. 91-116). Bristol: Multilingual Matters. https://doi.org/10.21832/9781853599286-008

Basterrechea, M., & Gallardo-del-Puerto, F. (2020). Language-related episodes and pair dynamics in primary school CLIL learners: A comparison between proficiency-matched and student-selected pairs. *Studies in Second Language Learning and Teaching*, *10*(3), 423-447. https://doi.org/10.14746/ssllt.2020.10.3.2

Basterrechea, M., & García Mayo, M. P. (2014). Dictogloss and the production of the English third person '-s' by CLIL and mainstream EFL learners: A comparative study. *International Journal of English Studies*, *14*(2), 77-98. https://doi.org/10.6018/j.177321

Benati, A. G. (2021). *Focus on form*. New York: Cambridge University Press. https://doi.org/10.1017/9781108762656

Bouffard, L. A., & Sarkar, M. (2008). Training 8-year-old French immersion students in metalinguistic analysis: An innovation in form-focused pedagogy. *Language Awareness*, *17*(1), 3-24. https://doi.org/10.2167/la424.0

Calzada, A. (2021). *Dictogloss in the primary school EFL classroom: Investigating the process, product and perceptions of collaborative writing.* [Unpublished doctoral dissertation, University of the Basque Country [UPV/EHU], Vitoria, Spain].

Calzada, A., & García Mayo, M. P. (2020). Child EFL grammar learning through a collaborative writing task. In W. Suzuki, & N. Storch (Eds.), *Languaging in language learning and teaching: A collection of empirical studies* (pp. 19-40). Amsterdam: John Benjamins. https://doi.org/10.1075/lllt.55.01cal

Calzada, A., & García Mayo, M. P. (2023). Do task repetition and pretask focus on form instruction impact collaborative writing performance? Evidence from young learners. In M. Li, & M. Zhang (Eds.), *L2 collaborative writing in diverse learning contexts* (pp. 80-108). Amsterdam: John Benjamins. https://doi.org/10.1075/lllt.59.04cal

Cambridge Assessment English. (2018). *Pre A1 Starters, A1 Movers and A2 Flyers: Sample papers for exams from 2018* (Vol. 1). Cambridge: Cambridge University Press.

Carroll, J. B., & Sapon, S. M. (2002). *Modern Language Aptitude Test - Elementary: Manual, 2002 edition*. Rockville, MD: Language Learning and Testing Foundation.

Cenoz, J. (2009). *Towards multilingual education: Basque educational research from an international perspective*. Bristol: Multilingual Matters. https://doi.org/10.21832/9781847691941

Collins, L., & Ruivivar, J. (2020). Form-focused instruction. In C. A. Chapelle (Ed.), *The encyclopedia of applied linguistics* (pp. 1-4). Hoboken, NJ: Wiley. https://doi.org/10.1002/9781405198431.wbeal0432.pub2

Collins, L., & Ruivivar, J. (2021). Research agenda: Researching grammar teaching and learning in the second language classroom. *Language Teaching*, *54*(3), 407-423. https://doi.org/10.1017/S0261444821000070

Collins, L., & White, J. (2014). The quantity and quality of language practice in typical interactive pair/group tasks. *TESL Canada Journal*, *31*(8), 47-67. https://doi.org/10.18806/tesl.v31i0.1186

DeKeyser, R. M. (2019). Aptitude-treatment interaction in second language learning: Introduction to the special issue. *Journal of Second Language Studies*, *2*(2), 165-168. https://doi.org/10.1075/jsls.00007.int

DeKeyser, R. (Ed.). (2021). *Aptitude-treatment interaction in second language learning*. Amsterdam, the Netherlands: John Benjamins. https://doi.org/10.1075/bct.116

Ellis, R. (2003). *Task-based language learning and teaching*. Oxford: Oxford University Press.

Ellis, R., & Barkhuizen, G. P. (2005). *Analysing learner language*. Oxford: Oxford University Press.

Ellis, R., Li, S., & Zhu, Y. (2019). The effects of pre-task explicit instruction on the performance of a focused task. *System*, *80*, 38-47. https://doi.org/10.1016/j.system.2018.10.004

Erlam, R. (2013). Effects of instruction on learners' acquisition of metalinguistic knowledge. In K. Roehr, & G. A. Gánem-Gutiérrez (Eds.), *The metalinguistic dimension in instructed second language learning* (pp. 71-94). London: Bloomsbury. https://doi.org/10.5040/9781472542137.ch-004

Ferman, S., & Karni, A. (2010). No childhood advantage in the acquisition of skill in using an artificial language rule. *PLoS ONE*, *5*(10), Article e13648. https://doi.org/10.1371/journal.pone.0013648

Fernández Dobao, A. (2012). Collaborative writing tasks in the L2 classroom: Comparing group, pair, and individual work. *Journal of Second Language Writing*, *21*(1), 40-58. https://doi.org/10.1016/j.jslw.2011.12.002

Fernández Dobao, A., & Blum, A. (2013). Collaborative writing in pairs and small groups: Learners' attitudes and perceptions. *System*, *41*(2), 365-378. https://doi.org/10.1016/j.system.2013.02.002

Fernández Dobao, A. (2020). Collaborative writing in mixed classes: What do heritage and second language learners think? *Foreign Language Annals*, *53*(1), 48-68. https://doi.org/10.1111/flan.12446

File, K. A., & Adams, R. (2010). Should vocabulary instruction be integrated or isolated? *TESOL Quarterly*, *44*(2), 222-249. https://doi.org/10.5054/tq.2010.219943

Frear, D., & Chiu, Y. (2015). The effect of focused and unfocused indirect written corrective feedback on EFL learners' accuracy in new pieces of writing. *System*, *53*, 24-34. https://doi.org/10.1016/j.system.2015.06.006

García Mayo, M. P. (2002). The effectiveness of two form-focused tasks in advanced EFL pedagogy. *International Journal of Applied Linguistics*, *12*(2), 156-175. https://doi.org/10.1111/1473-4192.t01-1-00029

García Mayo, M. P., & Imaz Agirre, A. (2019). Task modality and pair formation method: Their impact on patterns of interaction and LREs among EFL primary school children. *System*, *80*, 165-175. https://doi.org/10.1016/j.system.2018.11.011

Gesa, F., & Suárez, M. del M. (2022). Individual differences and young learners' L2 vocabulary development: The case of language aptitude and exposure to subtitled TV series. In V. De Wilde, & C. Goriot (Eds.), *Second language learning before adulthood: Individual differences in children and adolescents* (pp. 65-86). Berlin: De Gruyter. https://doi.org/10.1515/9783110743043-004

Goo, J., Granena, G., Yilmaz, Y., & Novella, M. (2015). Implicit and explicit instruction in L2 learning: Norris & Ortega (2000) revisited and updated. In P. Rebuschat (Ed.), *Implicit and explicit learning of languages* (pp. 443-482). Amsterdam: John Benjamins. https://doi.org/10.1075/sibil.48.18goo

Gorman, M., & Ellis, R. (2019). The relative effects of metalinguistic explanation and direct written corrective feedback on children's grammatical accuracy in new writing. *Language Teaching for Young Learners*, *1*(1), 57-81. https://doi.org/10.1075/ltyl.00005.gor

Gutiérrez, X. (2013). Metalinguistic knowledge in language-related episodes with covert metalinguistic activity. In K. Roehr, & G. A. Gánem-Gutiérrez (Eds.), *The metalinguistic dimension in instructed second language learning* (pp. 147-169). London: Bloomsbury. https://doi.org/10.5040/9781472542137.ch-007

Hakes, D. T. (1980). *The development of metalinguistic abilities in children*. Berlin: Springer. https://doi.org/10.1007/978-3-642-67761-8

Hanan, R. E. (2015). *The effectiveness of explicit grammar instruction for the young foreign language learner: A classroom-based experimental study*. [Doctoral dissertation, University of York, York, England].

Harley, B. (1998). The role of focus-on-form tasks in promoting child L2 acquisition. In C. J. Doughty, & J. Williams (Eds.), *Focus on form in classroom second language acquisition* (pp. 156-173). Cambridge: Cambridge University Press.

Harley, B. (1989). Functional grammar in French immersion: A classroom experiment. *Applied Linguistics*, *10*(3), 331-360. https://doi.org/10.1093/applin/10.3.331

Iglesias-Diéguez, K., & Martínez-Adrián, M. (in press). At the crossroad between form-focused instruction (FFI) and language-analytic ability: A study on metalinguistic explanations and the acquisition of explicit knowledge of the L3 English possessive determiners *his/her*. In M. P. García Mayo (Ed.), *Investigating attention to form and individual differences: Research with EFL children*. Cham: Springer.

Imaz Agirre, A., & García Mayo, M. P. (2013). Gender agreement in L3 English by Basque/Spanish bilinguals. *Linguistic Approaches to Bilingualism*, *3*(4), 415-447. https://doi.org/10.1075/lab.3.4.02ima

Izumi, S. (2003). Comprehension and production processes in second language learning: In search of the psycholinguistic rationale of the output hypothesis. *Applied Linguistics*, *24*(2), 168-196. https://doi.org/10.1093/applin/24.2.168

Johns, A. M. (2003). Genre and ESL/EFL composition instruction. In B. Kroll (Ed.), *Exploring the dynamics of second language writing* (pp. 195-217). Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9781139524810.014

Kachinske, I., & DeKeyser, R. (2019). The interaction between timing of explicit grammar explanation and individual differences in second language acquisition. *Journal of Second Language Studies*, *2*(2), 197-232. https://doi.org/10.1075/jsls.19003.kac

Kim, Y. (2008). The contribution of collaborative and individual tasks to the acquisition of L2 vocabulary. *The Modern Language Journal*, *92*(1), 114-130. https://doi.org/10.1111/j.1540-4781.2008.00690.x

Kiss, C., & Nikolov, M. (2005). Developing, piloting, and validating an instrument to measure young learners' aptitude. *Language Learning*, *55*(1), 99-150. https://doi.org/10.1111/j.0023-8333.2005.00291.x

Kopinska, M., & Azkarai, A. (2020). Exploring young EFL learners' motivation: Individual versus pair work on dictogloss tasks. *Studies in Second Language Learning and Teaching*, *10*(3), 607-630. https://doi.org/10.14746/ssllt.2020.10.3.10

Kuiken, F., & Vedder, I. (2012). Syntactic complexity, lexical variation and accuracy as a function of task complexity and proficiency level in L2 writing and speaking. In A. Housen, F. Kuiken, & I. Vedder (Eds.), *Dimensions of L2 performance and proficiency: Complexity, accuracy and fluency in SLA* (pp. 143-169). Amsterdam: John Benjamins. https://doi.org/10.1075/lllt.32.07kui

Lado, B., & Sanz, C. (2021). Feedback, aptitude, and multilingualism. In H. Nassaji, & E. Kartchava (Eds.), *The Cambridge handbook of corrective feedback in second language learning and teaching* (pp. 689-712). Cambridge: Cambridge University Press. https://doi.org/10.1017/9781108589789.033

Leeming, P., Aubrey, S., & Lambert, C. (2022). Collaborative pre-task planning processes and second language task performance. *RELC Journal*, *53*(3), 534-550. https://doi.org/10.1177/0033688220953225

Li, S. (2013). The interactions between the effects of implicit and explicit feedback and individual differences in language analytic ability and working memory. *The Modern Language Journal*, *97*(3), 634-654. https://doi.org/10.1111/j.1540-4781.2013.12030.x

Li, S. (2016). The construct validity of language aptitude: A meta-analysis. *Studies in Second Language Acquisition*, *38*(4), 801-842. https://doi.org/10.1017/S027226311500042X

Li, S., Ellis, R., & Kim, J. (2018). The influence of pre-task grammar instruction on L2 learning: An experimental study. *Studies in English Education*, *23*(4), 831-857. https://doi.org/10.22275/SEE.23.4.03

Li, S., Ellis, R., & Zhu, Y. (2019). The associations between cognitive ability and L2 development under five different instructional conditions. *Applied Psycholinguistics*, *40*(3), 693-722. https://doi.org/10.1017/S0142716418000796

Lichtman, K. (2021). What about fluency? Implicit vs. explicit training affects artificial mini-language production. *Applied Linguistics*, *42*(4), 668-691. https://doi.org/10.1093/applin/amaa054

Lim, W. L., & Jacobs, G. M. (2001). *An analysis of students' dyadic interaction on a dictogloss task*. ERIC (ED456649).

Loewen, S., & Sato, M. (2018). Interaction and instructed second language acquisition. *Language Teaching*, *51*(3), 285-329. https://doi.org/10.1017/S0261444818000125

Lyster, R. (2007). *Learning and teaching languages through content: A counterbalanced approach*. Amsterdam: John Benjamins. https://doi.org/10.1075/lllt.18

Lyster, R. (2015). Using form-focused tasks to integrate language across the immersion curriculum. *System*, *54*, 4-13. https://doi.org/10.1016/j.system.2014.09.022

Martínez-Adrián, M., & Gallardo-del-Puerto, F. (2021). Task modality and language-related episodes in young learners: An attempt to manage accuracy and editing. *Language Teaching Research, 28*(6), 2300-2325. https://doi.org/10.1177/13621688211052808

Michaud, G., & Ammar, A. (2023). Explicit instruction within a task: Before, during, or after? *Studies in Second Language Acquisition*, *45*(2), 442-460. https://doi.org/10.1017/S0272263122000316

Mitchell, R. (2000). Applied linguistics and evidence-based classroom practice: The case of foreign language grammar pedagogy. *Applied Linguistics*, *21*(3), 281-303. https://doi.org/10.1093/applin/21.3.281

Muñoz, C. (2014). The association between aptitude components and language skills in young learners. In M. Pawlak, & L. Aronin (Eds.), *Essential topics in applied linguistics and multilingualism: Studies in honor of David Singleton* (pp. 51-68). Cham: Springer. https://doi.org/10.1007/978-3-319-01414-2_4

Nassaji, H. (2016). Research timeline: Form-focused instruction and second language acquisition. *Language Teaching*, *49*(1), 35-62. https://doi.org/10.1017/S0261444815000403

Norris, J. M., & Ortega, L. (2001). Does type of instruction make a difference? Substantive findings from a meta-analytic review. *Language Learning*, *51*(s1), 157-213. https://doi.org/10.1111/j.1467-1770.2001.tb00017.x

Oliver, R., & Azkarai, A. (2017). Review of child second language acquisition (SLA): Examining theories and research. *Annual Review of Applied Linguistics*, *37*, 62-76. https://doi.org/10.1017/S0267190517000058

Oliver, R., & Bogachenko, T. (2019). Children and tasks. *Language Teaching Research Quarterly*, *12*, 48-65.

Oliver, R., Philp, J., & Duchesne, S. (2017). Children working it out together: A comparison of younger and older learners collaborating in task based interaction. *System*, *69*, 1-14. https://doi.org/10.1016/j.system.2017.08.001

Pawlak, M. (2021). Corrective feedback, developmental readiness, and language proficiency. In H. Nassaji, & E. Kartchava (Eds.), *The Cambridge handbook of corrective feedback in second language learning and teaching* (pp. 733-753). Cambridge: Cambridge University Press. https://doi.org/10.1017/9781108589789.035

Philp, J., Walter, S., & Basturkmen, H. (2010). Peer interaction in the foreign language classroom: What factors foster a focus on form? *Language Awareness*, *19*(4), 261-279. https://doi.org/10.1080/09658416.2010.516831

Pica, T. (1983). Adult acquisition of English as a Second Language under different conditions of exposure. *Language Learning*, *33*(4), 465-497. https://doi.org/10.1111/j.1467-1770.1983.tb00945.x

Plonsky, L., & Oswald, F. L. (2014). How big is "big"? Interpreting effect sizes in L2 research. *Language Learning*, *64*(4), 878-912. https://doi.org/10.1111/lang.12079

Quinn, P. (2014). *Delayed versus immediate corrective feedback on orally produced passive errors in English*. [Doctoral dissertation, University of Toronto, Toronto, Canada].

Roehr, K. (2008). Metalinguistic knowledge and language ability in university-level L2 learners. *Applied Linguistics*, *29*(2), 173-199. https://doi.org/10.1093/applin/amm037

Roehr-Brackin, K. (2018). *Metalinguistic awareness and second language acquisition*. New York: Routledge. https://doi.org/10.4324/9781315661001

Roehr-Brackin, K. (2022). Investigating explicit and implicit L2 knowledge and learning: Replications of Erlam (2005) and Roehr-Brackin and Tellier (2019). *Language Teaching*, *55*(2), 271-283. https://doi.org/10.1017/S0261444820000415

Roehr-Brackin, K. (2024). Explicit and implicit knowledge and learning of an additional language: A research agenda. *Language Teaching*, *57*(1), 68-86. https://doi.org/10.1017/S026144482200026X

Roehr-Brackin, K., & Tellier, A. (2019). The role of language-analytic ability in children's instructed second language learning. *Studies in Second Language Acquisition*, *41*(5), 1111-1131. https://doi.org/10.1017/S0272263119000214

Rosa-González, D. (2011). *The effects of aptitude and motivation on the acquisition of EFL in young learners.* [Master's thesis, University of Barcelona, Barcelona, Spain].

Serrano, R. (2011). From metalinguistic instruction to metalinguistic knowledge, and from metalinguistic knowledge to performance in error correction and oral production tasks. *Language Awareness*, *20*(1), 1-16. https://doi.org/10.1080/09658416.2010.529911

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

Skehan, P. (1998). *A cognitive approach to language learning*. Oxford: Oxford University Press.

Skehan, P. (2002). Theorising and updating aptitude. In P. Robinson (Ed.), *Individual differences and instructed language learning* (pp. 69-93). Amsterdam: John Benjamins. https://doi.org/10.1075/lllt.2.06ske

Spada, N. (1997). Form-focussed instruction and second language acquisition: A review of classroom and laboratory research. *Language Teaching*, *30*(2), 73-87. https://doi.org/10.1017/S0261444800012799

Spada, N. (2023). Form-focused instruction. In C. A. Chapelle (Ed.), *The encyclopedia of applied linguistics* (pp. 1-4). Hoboken: Wiley. https://doi.org/10.1002/9781405198431.wbeal0432.pub3

Stansfield, C. W., & Reed, D. J. (2005). *Modern Language Aptitude Test - Elementary: Spanish version, 2005 edition*. Rockville, MD: Language Learning and Testing Foundation.

Storch, N. (1999). Are two heads better than one? Pair work and grammatical accuracy. *System*, *27*(3), 363-374. https://doi.org/10.1016/S0346-251X(99)00031-7

Storch, N. (2016). Collaborative writing. In R. M. Manchón, & P. K. Matsuda (Eds.), *Handbook of second and foreign language writing* (pp. 387-406). Berlin: De Gruyter. https://doi.org/10.1515/9781614511335-021

Suzuki, Y. (2019). Individualization of practice distribution in second language grammar learning: The role of metalinguistic rule rehearsal ability and working memory capacity. *Journal of Second Language Studies*, *2*(2), 169-196. https://doi.org/10.1075/jsls.18023.suz

Suzuki, Y. (2022). Individual difference factors for second language grammar. In S. Li, P. Hiver, & M. Papi (Eds.), *The Routledge handbook of second language acquisition and individual differences* (pp. 294-309). New York: Routledge. https://doi.org/10.4324/9781003270546-25

Suzuki, Y., & DeKeyser, R. (2017). The interface of explicit and implicit knowledge in a second language: Insights from individual differences in cognitive aptitudes. *Language Learning*, *67*(4), 747-790. https://doi.org/10.1111/lang.12241

Suzuki, Y., Yokosawa, S., & Aline, D. (2022). The role of working memory in blocked and interleaved grammar practice: Proceduralization of L2 syntax. *Language Teaching Research*, *26*(4), 671-695. https://doi.org/10.1177/1362168820913985

Swain, M. (1995). Three functions of output in second language learning. In G. Cook, & B. Seidlhofer (Eds.), *Principle and practice in applied linguistics: Studies in honour of H. G. Widdowson* (pp. 125-144). Oxford: Oxford University Press.

Swain, M. (1997). Collaborative dialogue: Its contribution to second language learning. *Revista Canaria de Estudios Ingleses*, *34*, 115-132.

Swain, M., & Lapkin, S. (1995). Problems in output and the cognitive processes they generate: A step towards second language learning. *Applied Linguistics*, *16*(3), 371-391. https://doi.org/10.1093/applin/16.3.371

Tellier, A., & Roehr-Brackin, K. (2013). The development of language learning aptitude and metalinguistic awareness in primary-school children: A classroom study. *Essex Research Reports in Linguistics*, *62*(1), 1-28.

The Douglas Fir Group. (2016). A transdisciplinary framework for SLA in a multilingual world. *The Modern Language Journal*, *100*(S1), 19-47. https://doi.org/10.1111/modl.12301

Thepseenu, B., & Roehr, K. (2013). University-level learners' beliefs about metalinguistic knowledge. In K. Roehr, & G. A. Gánem-Gutiérrez (Eds.), *The metalinguistic dimension in instructed second language learning* (pp. 95-118). London: Bloomsbury. https://doi.org/10.5040/9781472542137.ch-005

Vraciu, A. (2020). Production accuracy of verb morphology in early EFL: Does primary school CLIL make a difference? *Linx*, *81*, 1-26. https://doi.org/10.4000/linx.7112

Wajnryb, R. (1990). *Resource book for teachers: Grammar dictation*. Oxford, England: Oxford University Press.

White, J., Muñoz, C., & Collins, L. (2007). The *his/her* challenge: Making progress in a 'regular' L2 programme. *Language Awareness*, *16*(4), 278-299. https://doi.org/10.2167/la419.0

White, J., & Ranta, L. (2002). Examining the interface between metalinguistic task performance and oral production in a Second Language. *Language Awareness*, *11*(4), 259-290. https://doi.org/10.1080/09658410208667060

White, L., Spada, N., Lightbown, P. M., & Ranta, L. (1991). Input enhancement and L2 question formation. *Applied Linguistics*, *12*(4), 416-432. https://doi.org/10.1093/applin/12.4.416

Xu, J., & Li, C. (2021). The effects of the timing of form-focused instruction on EFL learners' learning of difficult and easy grammatical features: A comparative study. *System*, *101*, Article 102612. https://doi.org/10.1016/j.system.2021.102612

Zhang, M., & Plonsky, L. (2020). Collaborative writing in face-to-face settings: A substantive and methodological review. *Journal of Second Language Writing*, *49*, Article 100753. https://doi.org/10.1016/j.jslw.2020.100753

Zimmerman, D. W., & Williams, R. H. (1982). Gain scores in research can be highly reliable. *Journal of Educational Measurement*, *19*(2), 149-154. https://doi.org/10.1111/j.1745-3984.1982.tb00124.x

Appendix 1 (Dictogloss 2)

Victor and Janire's family is on holiday in France. Victor speaks French very well. His father is French and his mum also speaks French. For Janire, Victor's sister, it is difficult to speak French. She constantly asks her brother to help her with the language. In the mornings, Victor goes surfing with his uncle and some new friends. Janire prefers riding horses so she rides with her aunt every day. In the afternoons, Victor likes to fish with his grandmother in their new boat. Janire prefers driving the boat with her mum. And in the evenings, her grandfather prepares a nice dinner in the boat for everyone!

[1] As conceived by Roehr (2008), metalinguistic knowledge can be defined as "a learner's explicit knowledge about the syntactic, morphological, lexical, phonological, and pragmatic features of the L2" (p. 179). It can be measured through metalinguistic tasks (e.g. grammaticality judgements, description and explanation tasks). It has also been investigated by examining learners' interaction while performing collaborative tasks in a L2 (operationalised through LRE production and accurate use of the target features selected) (Gutiérrez, 2013; Thepseenu & Roehr, 2013). As part of a bigger project on the effect of different forms of FFI treatments after dictogloss tasks on possessive determiners *his/her* and the third person singular *-s* morpheme (ME for the former and written corrective feedback for the latter), the present study will tackle the dimension of accuracy in the use of possessive determiners in written production, while the performance of a metalinguistic task is reported in Iglesias-Diéguez & Martínez-Adrián (in press) and LRE production will be explored in a forthcoming investigation.

[2] Note that this measure is also being used in the case of the studies within the same project examining FFI in the form of written corrective feedback and the third person singular *-s* morpheme.

[3] Notice that, as in previous studies analysing production data, the percentages obtained for each participant rely on a different number of OCs produced since these cannot be controlled for and depend on students' own production (e.g. Calzada & García Mayo, 2023).

[4] Note also that other focused tasks did not work as planned in other investigations with the possessive determiners *his/her* (Collins & White, 2014).

[5] The same dictogloss tasks have also proved effective in another forthcoming study conducted with a different sample in which both an experimental group receiving pre-task grammar instruction and a task-only group significantly improved from a pre- to a post-test phase.

[6] As noted by an anonymous reviewer, bilingual children, such as those that have taken part in this study, may have greater metalinguistic awareness than monolingual children. However, this is unlikely to have had any impact on the better results obtained by the group with ME as participants in both groups are bilingual. Note also that the blurring effect observed in this study has also been observed with Japanese monolingual adults (Suzuki et al., 2022).

[7] Even if future investigations should control for similar levels of MLAT-ES, it is worth mentioning that in a study with the same participants in which an *individual* grammaticality judgement task was incorporated, the trend observed in the current study as regards the blurring effect of ME has also been attested (see Iglesias-Diéguez & Martínez-Adrián, in press).

[8] Qualitative methodologies would have added further validity to the study, but as part of a bigger project, there were additional instruments that already took much time.