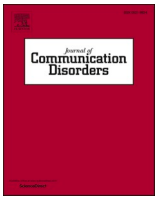
Journal of Communication Disorders 100 (2022) 106269

Contents lists available at ScienceDirect

Journal of Communication Disorders

journal homepage: www.elsevier.com/locate/jcomdis



Teaching foreign language grammar to primary-school children with developmental language disorder: A classroom-based intervention study

Elena Tribushinina [a,*], Geke Niemann [a], Joyce Meuwissen [b], Megan Mackaaij [a], Gabriëlla Lahdo [a]

[a] Utrecht University, Trans 10, Utrecht 3512 JK, the Netherlands
[b] Royal Kentalis, AB-dienst, Nijmeegsebaan 21a, Groesbeek GLD 6561 KE, the Netherlands

ARTICLE INFO

Keywords:
Developmental language disorder
Foreign language
Metalinguistic Intervention
Cross-linguistic awareness

ABSTRACT

Introduction: Children with developmental language disorder (DLD) start learning foreign languages, usually English as a foreign language (EFL), at an increasingly young age. However, current scholarship lacks crucial insights into how children with DLD respond to language learning in classroom settings and how they can be supported in doing so. The purpose of this early efficacy study is to determine whether a business-as-usual curriculum or a new teaching method tailored to the specific needs of pupils with DLD results in (greater) progress in the foreign language (English) and in the school language (Dutch).

Method: The participants were 75 pupils with DLD in the last three years of primary school, learning EFL in special education in the Netherlands. The intervention group (n=41) received 12 lessons following the CodeTaal approach, including metalinguistic instruction of grammar rules, explicit cross-linguistic contrasts and multimodal interaction with the material. The control group (n=34) received their regular English lessons. The study used a pre- to post-test design and compared the performance of the two groups on a Grammaticality Judgment Task (GJT) in English and a narrative task in both English and Dutch.

Results: Only the intervention group significantly improved in their ability to identify ungrammaticalities in English and generalised the learnt rules to new sentences. Although the performance on the GJT predicted accuracy of English narratives, neither group showed a significant decrease of error rates in English. In contrast, the accuracy of Dutch narratives showed improvement, but only in the intervention group. However, the effects were small and there was significant variability in responsiveness to the intervention.

Conclusion: We conclude that pupils with DLD are able to make progress in foreign language learning in a classroom setting if provided with adequate support.

1. Introduction

Around the globe children start learning foreign languages (FLs), usually English as a Foreign Language (EFL), at a very young age. In the recent decades, English has become a mandatory subject in primary schools in many countries across the world (Garton et al.,

* Corresponding author.
E-mail address: e.tribushinina@uu.nl (E. Tribushinina).

https://doi.org/10.1016/j.jcomdis.2022.106269
Received 23 February 2022; Received in revised form 30 August 2022; Accepted 26 September 2022
Available online 28 September 2022

2011; Gursoy et al., 2017; Rao & Yu, 2019). In the European Union, about 80% of pupils start English lessons already in primary school, and in several EU member states this number is close to 100% (Eurostat, 2019). The world-wide tendency towards an early onset of EFL instruction inevitably affects children with learning disabilities, and pupils with developmental language disorder (DLD) in particular (Tribushinina et al., 2020). Following the general trend, these children are also increasingly taught (E)FL from early on. For example, in the Netherlands, where the present study was conducted, EFL classes became mandatory in specialist primary schools for children with language disorders in 2012 (Thijs et al., 2011). All children with DLD start learning English in the 5th grade at the latest. In addition, many children with DLD, particularly those with a less severe disorder, are enrolled in mainstream schools where English lessons can already start in kindergarten (Thijs et al., 2011). However, we know very little about how children with DLD cope with FL learning and how they can be supported in doing so.

Research on second language (L2) acquisition by children with DLD has largely focussed on naturalistic settings, mainly in the countries/regions where English is the majority language. Research in such input-rich settings generally demonstrates that DLD is not an impediment to bilingualism (Paradis et al., 2021). In contrast, very little is known about how pupils with DLD manage with instructed L2 learning in a FL classroom, where exposure to the target language is limited to 30–90 min a week. This difference between naturalistic L2 acquisition and instructed FL learning is particularly important in the DLD context since the disorder is associated with deficits in procedural learning (Ullman & Pierpont, 2005), working memory (Ellis Weismer et al., 1999) and processing speed (Windsor, 2002). Due to these deficits, children with DLD need more exposure to the target language (and hence more time) than children with typical language development (TLD), even in input-rich naturalistic settings (Evans et al., 2009; Tomblin et al., 2007). The negative effects of the processing and procedural learning deficits should be particularly daunting in combination with very limited exposure provided by FL classrooms in school settings.

There are no published reports of intervention studies that developed FL teaching approaches catered to the specific needs of pupils with DLD, and there are only three published studies focussing on EFL learning in this group of children (Tribushinina et al., 2020, 2022a; Zoutenbier & Zwitserlood, 2019). These studies demonstrate that primary-school children with DLD have difficulty acquiring various EFL skills and learn more slowly than their peers with TLD. These converging results are particularly interesting because the studies were conducted in very different cultural and instructional contexts. Tribushinina et al. (2020) traced EFL progress of children with and without DLD in the first two years of primary education in Siberia. Their participants had very little exposure to English outside of the classroom, and children with DLD had even less extramural exposure than typically-developing pupils. The results demonstrated that there were no differences between pupils with and without DLD after one year of English lessons. However, children with TLD made significant progress in both vocabulary and grammar after 1.5 and 2 years of English lessons, whereas this was not the case for the DLD group. Another interesting finding was that L1 vocabulary and grammar predicted EFL vocabulary and grammar in typically-developing children, but in the DLD group such positive cross-language relationships were found only in the domain of vocabulary (cf. Tribushinina et al., 2022a). The authors suggest that DLD may selectively affect the mechanisms of positive cross-language transfer in the vulnerable domain of morphosyntax.

Zoutenbier and Zwitserlood (2019) targeted English skills in Dutch-speaking sixth graders with DLD. In the Netherlands, children usually have a lot of out-of-school exposure to English (e.g., all media are subtitled rather than dubbed). Yet, even in this favourable context pupils with DLD scored lower than age-related norms on listening, reading and vocabulary skills in English. One of the reasons Dutch-speaking primary-school children with DLD perform so poorly in English could be that the teaching approach pursued by Dutch primary schools might not be suitable for learners with DLD. Specialist education facilities for pupils with language disorders use the same EFL teaching methods as mainstream schools, and pupils with DLD are expected to learn English in the same way as pupils with TLD (Thijs et al., 2011). These teaching practices disregard recent research findings demonstrating that children with DLD have difficulty with FL learning (Tribushinina et al., 2020; Zoutenbier & Zwitserlood, 2019) and that their learning mechanisms are not the same as in peers with TLD (Lum et al., 2012; Ullman & Pierpont, 2005).

This paper reports a first attempt to develop an EFL teaching approach tailored to the specific needs of pupils with DLD. Our intervention, called *CodeTaal* (Code Language), was implemented in the final three grades of a specialist primary school in the Netherlands. The efficacy of the intervention is tested by comparing progress in the intervention group and a business-as-usual control group. This study brings together two research lines: (i) intervention research in language disorders and (ii) research on instructed L2 learning and FL pedagogy.

1.1. *Procedural learning disadvantages in DLD*

There is growing evidence that DLD involves difficulty with implicit learning. Children with DLD tend to perform worse than typically-developing peers on non-linguistic Serial Reaction Time tasks that involve unconscious detection of statistical regularities, such as patterns in the flow of visual stimuli (e.g., shapes) (see meta-analyses in Lammertink et al., 2017; Lum et al., 2014; Obeid et al., 2016). This difficulty is presumably caused by abnormal functioning of procedural memory that is responsible for implicit learning of cognitive routines and motor skills (Ullman & Pierpont, 2005). In the linguistic domain, procedural memory supports the acquisition of rule-based aspects of language (regular morphology, syntax), i.e., domains that are particularly problematic in DLD (Leonard, 2014). In contrast, declarative memory, which supports conscious learning of factual knowledge (what/where/when), is supposedly spared in DLD (Ullman & Pierpont, 2005). Declarative memory underlies the acquisition of idiosyncratic language elements (words, irregular morphology); these are the areas of relative strength in DLD (Leonard, 2014).

It is not entirely clear how the procedural learning deficit develops with age. In children with TLD, procedural memory is well-established early in life and attenuates with age, whereas declarative memory improves throughout childhood (Ullman, 2016). For DLD, some studies suggest that procedural learning remains problematic throughout childhood (Lammertink et al., 2017; Obeid et al.,

2016), but there is also evidence that the gap between children with and without DLD in procedural-learning tasks becomes smaller with age (Lum et al., 2014). The negative impact of the deficit may abate due to improving declarative memory and/or increasing compensation through declarative memory (Lukács et al., 2017).

It has been argued that, in children with TLD, declarative memory predicts vocabulary skills, whereas procedural memory is associated with grammar skills (Ullman, 2016). In contrast, in children with DLD, declarative memory appears to predict both vocabulary and grammar knowledge (Lum et al., 2012). However, more recent evidence regarding these group differences is rather inconclusive. For example, an experimental study and a meta-analysis reported by Lammertink et al. (2020) found no evidence for (or against) the differential relationship between procedural learning ability (in the visuo-spatial domain) and grammatical ability in children with and without DLD. This said, it has been shown on multiple occasions that children with DLD are likely to memorise regular forms (e.g., *opened*) through declarative memory instead of applying the relevant rules through procedural memory (e.g., Gopnik & Crago, 1991) and that they benefit from interventions drawing on explicit grammar rules that are also learnt through declarative memory (Ebbels, 2014).

1.2. Explicit teaching approaches and metalinguistic interventions

Primary-school children usually engage with FLs in a playful manner, for example, through songs and child rhymes. EFL instruction in Dutch primary-schools is skill-based: It mainly targets vocabulary (such as colour terms and days of the week) and basic conversational routines (e.g., talking about the weather or hobbies). The focus is on oral communication and on reading short texts. Writing and spelling practice is limited to some common words. Grammar rules are usually not taught, as communicative competence is deemed more important than knowledge of grammar (Rose, 2016; Thijs et al., 2011). The dominant teaching approach adopted in primary education in the Netherlands can thus be called *implicit* (see Hanan, 2015 for a discussion of a similar situation in the UK). Implicit language learning involves unconscious input processing without systematic attention to form and results in procedural knowledge of how to use language (Hulstijn, 2005). Implicit teaching does not include either rule explanation (deductive teaching) or tasks to search for regularities in the input (inductive teaching). Teachers may, however, incidentally draw the learners' attention to form if the communicative need arises. Corrective feedback is usually provided in the form of recasts that do not disrupt the flow of natural communication (Piggott, 2019). In contrast, explicit language learning is an intentional process, ranging from conscious rule detection to learning through metalinguistic explanation. Explicit grammar teaching is more common in secondary schools, even though it appears to be equally effective with younger children (Lichtman, 2016).

Comparison of the effectiveness of explicit and implicit grammar teaching approaches has been a topical issue in L2 research for several decades. Explicit approaches are usually associated with greater learning gains in limited-input settings (e.g., Norris & Ortega, 2000; Spada & Tomita, 2010), but the outcome measures tend to be biased towards explicit knowledge (Piggott, 2019). Although it is possible to acquire L2 morphosyntax without explicit rule teaching (Rebuschat & Williams, 2012), limited classroom time makes it difficult to learn (E)FL grammar solely through procedural memory (implicitly), even for children with TLD. Explicit approaches appear particularly beneficial to low-aptitude learners (Erlam, 2005) and students with weaker procedural memory (Tagarelli et al., 2016). High-aptitude learners are good at detecting statistical regularities and are therefore less dependent on metalinguistic instruction. When it comes to low-aptitude learners, who are less good at noticing patterns in the input, explicit teaching approaches help them by drawing their attention to regularities that may otherwise go unnoticed (Benson & DeKeyser, 2019; Hwu et al., 2014).

In view of the growing body of evidence that DLD is associated with procedural learning deficits (see above), it is plausible that FL learners with DLD will have even more trouble learning FLs implicitly than peers with TLD. It is also likely that FL learning success of pupils with DLD crucially depends on the presence of explicit metalinguistic explanations. Metalinguistic interventions explicitly teaching children morphosyntax have proven effective for remedying and enhancing L1 development of children with DLD in speech and language therapy (see reviews in Balthazar et al., 2020; Ebbels, 2014; Frizelle et al., 2021). Such approaches explicitly draw the learners' attention to morphosyntactic regularities and promote learning through declarative memory that is supposedly spared in DLD.

Metalinguistic interventions for children with DLD often involve visual support, such as colour coding of argument structure in *Colourful Semantics* (Bryan, 1997) or a combination of shapes and colours in *Shape Coding*, where colours code parts of speech, shapes indicate phrasal structures and arrows capture tense and aspect (Ebbels, 2007). Such approaches have been shown to benefit various aspects of L1 grammar, such as verb morphology (Calder et al., 2018, 2021; Ebbels, 2007; Seeff-Gabriel et al., 2012; Tobin & Ebbels, 2019), verb argument structure (Bolderson et al., 2011; Ebbels et al., 2007; Levy & Friedmann, 2009; Spooner, 2002), question formation (Ebbels, 2007), dative structures (Ebbels, 2007), subordinate clauses (Balthazar & Scott, 2018), passives and wh-questions (Ebbels & Van der Lely, 2001). Sometimes a tactile/kinaesthetic dimension is added to enhance levels of active participation and engagement with the material, as in the *MetaTaal* approach, using LEGO bricks to support production and comprehension of relative clauses in Dutch (Zwitserlood et al., 2015).

Although most metalinguistic interventions targeted older secondary-school children, evidence is accumulating that such approaches are also suitable for younger children (Bolderson et al., 2011; Calder et al., 2018, 2021; Seeff-Gabriel et al., 2012). The majority of such interventions have been tested in one-to-one sessions with a speech and language therapist, but there is also some evidence that therapy delivered to small groups of children (Tobin & Ebbels, 2019) or even to whole classes (Ebbels, 2007; Hirschman, 2000) can be effective. In fact, as suggested by Hirschman (2000), metalinguistic interventions are particularly suitable for working with larger groups because they are less labour-intensive than implicit techniques such as recasting. In our study, the intervention was implemented to whole classes during the time they would otherwise have their regular English lessons.

1.3. Interventions and teaching approaches raising cross-linguistic awareness

The idea that target language use should be maximised in FL classrooms and L1 use should be avoided has dominated FL pedagogy for decades (Littlewood & Yu, 2011). Following this general trend, national guidelines for FL curricula in the Netherlands stipulate that children should be immersed in the target language and that English should be used both as the target language and the medium of instruction in EFL lessons (Haamberg et al., 2008). However, this monolingual orientation violates an important condition for effective learning, which involves constructing new understandings on the basis of the already acquired understandings and experiences (Donovan & Bransford, 2005). Critically, L1 knowledge is not an impediment to FL learning, rather it is "a building block for the second language providing a scaffold for its development" (Wigglesworth, 2002, p. 19). For instance, if a child already knows the word *optimistisch* in their L1 Dutch, it should be easy to learn the meaning of the English cognate *optimistic*. Similarly, understanding that L1 Dutch verbs can be regular (e.g., *open-opende*) and irregular (e.g., *hang-hing*) will facilitate the acquisition of regular (e.g., *open-opened*) and irregular (e.g., *hang-hung*) past tense forms in L2 English (Blom & Paradis, 2015).

Classroom observation studies reveal that L2 learners rarely engage in spontaneous comparisons of L1 and L2 (Bell et al., 2020) and are often unaware of crucial cross-linguistic differences (Ammar et al., 2010; Bouffard & Sarkar, 2008). Systematic cross-language comparisons not only reinforce positive transfer, they also pre-empt negative transfer by showing what *cannot* be done in the L2, particularly when it comes to teaching grammatical features that display subtle but meaningful cross-linguistic differences. For example, Dutch learners of English often make word order errors because English and Dutch differ in the position of frequency adverbials (*Jan gaat vaak naar de film* vs. *John often goes to the movies*). Prior research shows that L2 learners are prone to accept both grammatical and ungrammatical L2 structures because they are not aware of how L1 and L2 differ, which results in transfer errors (Lightbown & Spada, 2000). Therefore, it has been suggested that L2/FL teachers should use pedagogical approaches making the learners aware of the L1-L2 form-meaning differences (Bell et al., 2020; Bouffard & Sarkar, 2008; Lightbown & Spada, 2000).

Intervention studies empirically testing these claims are still scarce (see McManus, 2022 for an overview). In a series of recent studies, McManus and Marsden (2017, 2018, 2019) aimed to determine whether explicit L1 instruction facilitates the learning of the French *Imparfait* by adult L1 English speakers. *Imparfait* expresses both past habituality and past ongoingness, whereas in English these meanings are mapped onto two different aspectual forms – Past Simple and Past Progressive, respectively. The results demonstrated that explicit instruction on L1 and L2 form-meaning mappings in tandem with comprehension practice in both languages led to greater learning gains compared to an L2-only approach and compared to an approach that included L1 practice but excluded explicit instruction about L1. McManus (2019) extended this line of research by demonstrating that explicit instruction about L1 and L2 leads to greater levels of awareness of form-meaning mappings, which in turn is associated with better performance in a range of comprehension and production tasks. It is concluded that the results support cognitive theories positing an important role of awareness in L2 learning (e.g., Tyler, 2012). Crucially, these findings demonstrate that awareness of L1-L2 mappings is more important than awareness of the L2 rule alone.

In a similar vein, Lucas (2020) investigated the effect of instruction drawing on explicit L1-L2 contrasts in the learning of English plurals by L1 Japanese adolescents. Plurals constitute a vulnerable feature in the L2 grammars of L1 Japanese speakers because, in isolating languages, the meaning of plurality is not marked by inflectional morphology but rather inferred from context (as in 'two apple'). The results revealed that exposure to online awareness-raising exercises led to gains in both error recognition and plural production.

To the best of our knowledge, only one study has applied a similar approach raising cross-linguistic awareness to teaching EFL to students with learning disabilities. Tribushinina et al. (2022b) explicitly taught EFL spelling to a group of Dutch-speaking students with dyslexia. Prior to introducing the target spelling rule in English, the teacher invited students to reflect on certain spelling regularities in L1 Dutch. After that, a similar rule in L2 English was introduced and explicitly compared to the rule in Dutch. Similarities and differences between the two languages were explicitly addressed in each lesson. For example, in both languages words cannot end in *v*, but Dutch and English solve this problem in different ways: In Dutch a word-final *v* becomes *f* (*dove-doof*), whereas in English the silent *e* is added (*nerve*). The progress in the intervention group was compared to that of a control group of dyslexic students attending their regular English lessons (without formal spelling instruction). The results showed that the intervention group made greater progress than the control group after only eight 20-min lessons and maintained the enhanced level of orthographic skills five weeks after the intervention.

We hypothesise that a teaching approach making L1-L2 differences and similarities explicit should also be beneficial to EFL learners with DLD. First, such teaching approaches are likely to facilitate positive cross-language transfer by activating L1 features that are similar to their L2/FL counterparts (like in the above example with regular past tense in English and Dutch). Even though very few studies addressed the ability of children with DLD to capitalise on their L1 knowledge in L2/FL learning, there are indications in the literature that children with DLD have difficulty spontaneously using positive transfer from their L1, particularly in the domain of morphosyntax (Blom & Paradis, 2015; Tribushinina et al., 2020). Second, cross-linguistic grammar instruction is likely to enhance morpho-syntactic awareness (Bouffard & Sarkar, 2008), which is predictive of L2 learning success (Fumero & Tibi, 2020) but is known to be weak in DLD (Kamhi & Koenig, 1985). Finally, there is evidence (obtained in naturalistic and immersive settings) that supporting L1 development creates a stronger foundation for developing L2 skills in children with DLD (Armon-Lotem et al., 2021; Dam et al., 2020; Perozzi & Sanchez, 1992). For example, Perozzi and Sanchez (1992) report a study in which one group of Spanish-English bilinguals first received instruction about prepositions in Spanish (L1) and then in English (L2), whereas another group of Spanish-English bilinguals only received instruction in the L2. Interestingly, the former group learnt English prepositions and pronouns faster, and they learnt them in both languages rather than just in English. It remains to be seen whether these results obtained in immersive settings can be generalised to FL learning in the classroom and to the domain of morphosyntax.

1.4. Grammaticality judgment tasks as outcome measures

Both research traditions relevant to the present study often use grammaticality judgment tasks (GJTs) but for different purposes. In L2 research, the ability to recognise errors is often used as an outcome measure in intervention studies comparing the effectiveness of explicit and implicit teaching approaches (e.g., Lucas, 2020; Rebuschat & Williams, 2012; Tagarelli et al., 2016). Crucially, grammatical and ungrammatical sentences appear to measure different constructs, with grammatical sentences tapping more into implicit knowledge and ungrammatical sentences drawing on explicit knowledge (Ellis, 2005).

In research on language disorders, GJTs are usually used to study language development and language profiles of children with DLD. Grammaticality judgements have been shown to be clinical markers of DLD (Rice et al., 1999). One common criticism is that GJTs require higher levels of metalinguistic awareness, which is problematic in DLD (Kamhi & Koenig, 1985). This notwithstanding, performance in GJTs has been shown to be a strong predictor of language production: Children are more likely to notice errors that they do not make in their own speech and less likely to detect errors that they themselves make (Rice et al., 1999; Smith-Lock, 1995). GJTs are not commonly used as outcome measures in clinical intervention studies. However, recent work by Calder and colleagues (2018, 2020, 2021) demonstrates that these tasks can be informative about the efficacy of interventions targeting morphosyntax and that they can be successfully used with children as young as age 6.

The present study will use a GJT to compare progress in the intervention group and the business-as-usual control group. We will also collect production data to determine whether GJT performance predicts accuracy in language production. Since auditorily presented GJTs impose a working memory demand (McDonald, 2008) and are also constrained by phonological sensitivity (Gottardo et al., 1996), we will present the stimuli both auditorily and visually.

2. The present study

This paper reports an exploratory study testing the efficacy of CodeTaal, a new EFL teaching approach developed specifically for primary-school pupils with DLD. Informed by the prior research discussed above, CodeTaal involves metalinguistic explanations, explicit cross-linguistic comparisons and multimodal engagement with the material. The following research questions are addressed:

1 Does the CodeTaal approach lead to greater gains in the ability to detect ungrammaticality in English?
   a And if so, are the gains generalised to non-treated items (of the treated grammatical categories) and to non-treated grammatical categories?
2 Does the CodeTaal approach lead to greater accuracy gains in semi-spontaneous language production in the target language (English) and in the school language (Dutch)?
3 Does performance on the GJT predict accuracy in semi-spontaneous language production?

Assuming that EFL learning by children with DLD is compromised by both procedural learning deficits (Ullman & Pierpont, 2005) and reduced capacity to instantiate positive cross-language transfer (Blom & Paradis, 2015; Tribushinina et al., 2020), we expect that a metalinguistic approach explicitly comparing L1 and L2 grammar should be beneficial to EFL learning. We expect that gains will be achieved in both grammaticality judgements and in production. We further hypothesise that the progress in detecting ungrammaticality will extend to new items/sentences not treated in the intervention (e.g., Bolderson et al., 2011; Calder et al., 2021; Ebbels et al., 2007; Finestack, 2018; Levy & Friedmann, 2009) but not to grammatical categories not treated in the intervention (e.g., Calder et al., 2021; Ebbels et al., 2007). Explicit cross-language comparisons may have a concurrent positive effect on the development of grammatical accuracy in the majority language (Dutch). Based on studies using GJTs with L1 learners of English (Rice et al., 1999; Smith-Lock, 1995), we hypothesise that performance on the GJT will be predictive of language production in EFL learners as well.

3. Method

This research was approved by the ethics assessment committee of the Faculty of Humanities at Utrecht University. Informed consent was obtained from the parents/guardians of the participants and from the participating schools.

3.1. Research design

The study employed a quasi-experimental pre- to post-test design, with the intervention group (IG) receiving 12 weekly lessons following the CodeTaal approach and the business-as-usual control group (CG) receiving their regular English lessons. The control group had been originally planned to be a waiting control group that would start the intervention in the spring term of the same academic year. However, this intervention had to be discontinued due to a COVID-19 lockdown. Two schools participated in this study: All intervention classes came from School 1 and all control classes came from School 2. For logistic reasons, randomisation of schools or classes was not possible.

3.2. Participants

Seventy-five children with DLD, aged 8;11–13;8, participated in this study (13 female). They had been independently diagnosed for DLD following a standardised protocol (Stichting Siméa, 2014), which requires an overall score of at least 2 *SD* below age-appropriate

norms on a standardised Dutch language test (usually CELF-4-NL) or scores of at least 1.5 *SD* below the age-appropriate mean score on at least two of the four subscales of a standardised language test. A hearing impairment and intellectual disability constituted exclusion criteria.

The participants attended specialist primary schools in the Netherlands (cluster-2 schools for children with language disorders and hearing impairments). Forty-one participants (8 female) were included in the IG and 34 children (5 female) were included in the CG. All participants spoke Dutch, and 19 participants also spoke another language at home (11 in the IG and 8 in the CG). There were no differences between the IG and the CG in age and proficiency in Dutch (Table 1).[1]

The participants attended one of the three final grades of primary education (Grade 4–6). These grades were selected because the participating schools taught English from Grade 4 onwards. The participants of the IG were drawn from three different classes (one Grade 4, one Grade 5 and one Grade 6 class) in the same primary school. The participants of the CG came from four different classes (two Grade 4 classes, one Grade 5 class and one Grade 6 class) recruited from a different primary school. The two schools were located in different cities but had a similar structure and educational philosophy; both schools belonged to the same overarching organisation providing specialised remedial education to children with language disorders. All participants received one 45-min English lesson a week. To keep time-on-task comparable across the two groups, the intervention was administered during the regular English lessons. The intervention started in the autumn term of the school year 2019-2020. So for the 4th graders the intervention was part of their first English lesson, whereas 5th and 6th graders had received prior formal instruction in English for one and two years, respectively.

3.3. Intervention

The CodeTaal intervention was inspired by multimodal metalinguistic approaches, including Montessori language symbols (Montessori, 1912), *Shape Coding* (Ebbels, 2007) and *MetaTaal* (Zwitserlood et al., 2015). The main principles of our approach are:

(i) *Metalinguistic*: The pupils are provided with explicit grammar instruction and learn linguistic terminology.
(ii) *Contrastive*: Knowledge of the corresponding Dutch structures is activated and made explicit prior to introducing the target rule in English. Similarities and differences between Dutch and English are explicitly addressed. To emphasise the cross-linguistic contrasts (e.g., placement of time adverbials before verbs in English and after verbs in Dutch), sentences are coded on a white Velcro tape for Dutch and on a black Velcro tape for English.
(iii) *Multimodal*: Teaching and learning hinge on visual support (written sentences; colour-coded shapes; white *vs.* black Velcro tape), auditory information (Dutch and English sentences pronounced by the teacher and orally discussed grammar rules), as well as kinaesthetic interaction with the materials (attaching coloured shapes to the Velcro tape).

In the CodeTaal approach colour-coded shapes are used for grammar instruction, in which each colour and shape represent a unique part of speech (e.g., a dark-green rectangle for nouns; a light-green square for personal pronouns) or morpheme (e.g., a red circle representing the verb stem; triangles of different colours coding verb endings). The shapes for function words and morphemes are smaller than the shapes for content words. These colour-coded shapes are similar to the Montessori language symbols often used in Dutch primary schools. An important difference is that our coding scheme also contains morphemes. Given that omission of (verb) inflections is a clinical marker of DLD in both English (Rice, 2000) and Dutch (Blom et al., 2013), we reckoned it crucial that the pupils would systematically attach the morpheme shapes to the verb stem, thereby making morphemes more salient. For example, a light blue triangle represents the third person singular *-s*, which only exists in English, whilst the corresponding Dutch morpheme *-t* is represented by a dark blue triangle. By using the same shape but two different shades of blue we emphasised that the two morphemes have the same function in both languages even though they differ in phonetic realization.

Each lesson started with a vocabulary learning activity that took about 10 min. Each treated word was spoken out loud in English by the teacher and the students were asked to provide the Dutch translation. In order to help the learners guess the meaning of the new words, pictures corresponding to all the words treated in the lesson were shown on a PowerPoint slide. After that, the pupils were asked to decide whether the words in the two languages were siblings (cognates). Sample lesson slides can be found in the Supplementary Materials.

The remaining part of the lesson (approximately 35 min) was devoted to the CodeTaal intervention. First, the target rule was explained or reiterated by the teacher. Then, the pupils practiced with "building" sentences in CodeTaal in pairs. Each pair had a printed overview of all of the symbols (with their meanings and examples), a box with coloured shapes sorted by category, and two Velcro strips – a white strip for Dutch and a black one for English. The pairs started by coding a Dutch sentence and then proceeded with coding its English counterpart. After that, the coded sentences were discussed with the whole class, while the teacher was coding the sentences by attaching large magnetic shapes to the whiteboard. The coded sequences were also presented on a PowerPoint slide: The Dutch sequence was always presented above the corresponding English sequence to make structural similarities and differences more salient (as in Fig. 1). Cross-linguistic similarities and differences were explicitly discussed. For example, the word order and verb morphology in the example sentences in Fig. 1 are identical in Dutch and English. The only difference is that Dutch uses *-t* as a third person singular morpheme, whereas English uses *-s*. Afterwards, that the pupils moved to a new pair of sentences. Sixth graders on

[1] The test scores are based on the yearly tests conducted by speech-language therapists of the participating schools. Recent PPVT scores were not available for 9 participants (all in the IG), and up-to-date CELF scores were not available for 9 participants (1 in the CG and 8 in the IG). The PPVT and the CELF scores were not normally distributed. Therefore, we report non-parametric statistics for these tests.

Table 1
Participant characteristics.

	Intervention		Control		Group comparisons
	Mean	SD	Mean	SD	
Age (months)	133.1	13.4	131.9	12.0	$t(73) = 0.40$, $p = .691$, $d = -0.09$, 95% CI [-0.55, -0.36]
CELF core score	66.3	9.73	66.4	11.6	$U = 507.00$, $z = -.485$, $p = .628$, $d = -0.36$, 95% CI [-0.85, -0.12]
PPVT	84.1	10.2	79.0	16.8	$U = 453.50$, $z = -1.162$, $p = .245$, $d = 0.01$, 95% CI [-0.47, - 0.49]

average coded three pairs of (Dutch-English) sentences per lesson, whereas 4th and 5th graders were a bit slower and could only practise one or two sentence pairs.

There were 12 intervention sessions taught over the span of 14 weeks. The intervention was delivered by a remedial teacher with expertise in DLD, who was not the regular teacher of the three intervention classes, but was familiar with the school and the pupils. Training and ongoing support was provided by the second author (researcher and certified English teacher) who had weekly meetings with the intervention teacher and was also present in the lessons to ensure treatment fidelity. The intervention procedure had been previously piloted in three one-to-one sessions with similar-aged children with DLD (none of them participating in the present study). The grammatical structures targeted in the intervention included third person singular in the Present Simple, conjugation of *to be* (copula) and *to have* (main verb), noun plurals, determiners and word order in declarative sentences.

3.4. English lessons in the control group

Since the goal of this research was to compare the effectiveness of our new teaching approach to that of the approach currently used in (specialist) primary schools in the Netherlands, the CG received their regular English lessons. The control classes used the coursebook *Groove.me* (produced by Blink), a method that is commonly used in primary schools in the Netherlands. This method teaches English based on pop-songs. The exercises focused on vocabulary, grammar, reading, writing, and listening skills and were always coupled to a specific song. Grammar was implicitly integrated in exercises focusing on writing, reading and speaking. No metalinguistic explanations and no comparisons between English and Dutch grammar were provided, as established through classroom observations, textbook analysis, and teacher interviews. As is common practice in Dutch primary schools, the English lessons were taught by the class teacher; the four participating teachers had a similar background and training as the remedial teacher delivering the intervention. The CG was matched with the IG on duration (45 min) and intensity (once a week) of English classes.

3.5. Test instruments and procedures

A GJT was used to compare pre-post intervention progress in the ability to detect ungrammaticality (in English). A narrative task was used to test grammatical accuracy in semi-spontaneous speech (in both English and Dutch). The tasks were administered in different sessions in the same week and their order was counterbalanced across participants. The tasks were administered by trained research assistants, all Dutch native speakers with university level training in English. The children were not familiar with the experimenters. The participants were tested individually in a quiet room in their school. The tests were administered twice: a week before the intervention started (pre-test) and a week after the intervention finished (post-test).

3.5.1. Grammaticality Judgment Task (GJT)

Materials. The GJT contained 54 English sentences, including 27 sentences with a grammatical error and 27 grammatically correct sentences. There were 18 fillers and 36 sentences representing the grammatical categories targeted in the intervention: third person singular present tense, determiners, noun plurals, the verbs *to be* and *to have*, and word order, with sets of 6 sentences targeting each targeted structure (half ungrammatical). Each set of 6 included 2 sentences that had been literally treated in the intervention (shape-coded in CodeTaal), as well as 4 sentences that were new but represented the same grammatical categories. Including both new and treated exemplars of the target structures allowed us to determine the ability of the participants to generalise the learnt rule to new instances. The 18 fillers targeted past simple, Present Progressive and Present Perfect (3 grammatical and 3 ungrammatical sentences per targeted structure). All stimuli were presented in a pseudo-randomised order. The test started with three practice trials, of which only the first contained a grammatically correct sentence. The test had been previously piloted with two children who did not participate in the main study. The test materials can be found in the Supplementary Materials.

Procedure. The GJT was administered using PowerPoint. The task started by introducing a bear who was learning English and sometimes made mistakes. The child was asked to help the bear by pointing out the errors in his sentences. The sentences were presented both visually and orally. The participants saw 57 slides (3 practice trials and 54 test trials) containing an image of the bear as well as the written version of the target sentence and heard an audio recording of the sentence. Although each audio recording was only played once, the participants could take as long as necessary to read the sentence and provide an answer. For each sentence, the participants were asked to indicate whether the sentence was grammatically correct or not. If the child said it was incorrect, s/he was asked to indicate what was wrong and/or how the sentence should be corrected. The participants were only given feedback on the three practice trials. It took the participants about 20 min to complete the test. They were allowed to take four short breaks during the experiment. The responses were audio-recorded and later transcribed for further coding.

Scoring. Each response was scored as 1 (correct) or 0 (incorrect, no response). Sentences appropriately judged as grammatical

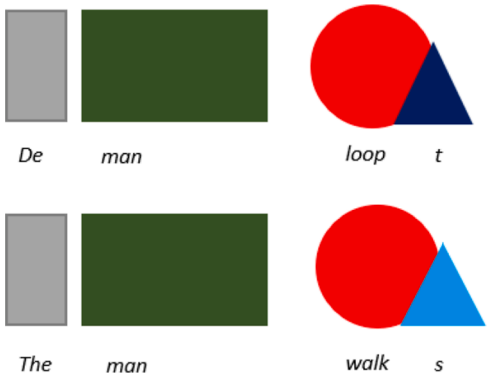


Fig. 1. Dutch and English equivalent sentences coded.

yielded one point. Sentences correctly marked as ungrammatical were only awarded one point if the participant managed to identify, explain, or correct the error. If a correction contained a mistake but indicated the participant understood the type of error, a point was still awarded. By way of illustration, consider the following two examples (the answers are translated here for convenience).

(1) *Yesterday I go to school.*
Participant's correction: It should be *goed to school* because *yesterday* means *gisteren*[2] and that is already past.

(2) *The boy drive in the yellow cars.*
Participant's correction: *Cars* should be *car* because it says *boy* and one boy can only drive in one car at the same time.

Although *goed* in example (1) is grammatically incorrect, the participant successfully determined the tense of this sentence and located the error. Hence, this answer was given 1 point. In contrast, the answer in (2) was not awarded a point. Although the participant understood the error in number, they did not identify the subject-verb agreement error (*boy drive*).

The coding was performed by the last author, in consultation with the first author. All cases of disagreement were discussed and resolved by consensus.

3.5.2. Narrative task

Materials. The narratives were elicited by means of the Multilingual Assessment Instrument for Narratives (Litmus-MAIN) (Gagarina et al., 2012). MAIN was developed for research with multilingual children and comprises four parallel sets of six coloured pictures (Cat, Dog, Baby Birds and Baby Goats), representing four stories with identical story and episodic structure. Even though the focus of the present study is on EFL proficiency, we decided to elicit narratives in both languages. Since the participants had only just started learning English and their overall proficiency was low, we reckoned that it would be easier if they could first practice in their stronger language (Dutch) and after that tell a similar story in English. In addition, the Dutch narratives could be used to pinpoint possible effects of the intervention on Dutch grammar skills. Two different MAIN stories were used to elicit the narratives: the Cat Story and the Dog Story. If the child told the Cat Story in Dutch during the pre-test, then they produced the Cat Story in English during the post-test and vice versa.

Procedure. The MAIN stories were elicited following the procedure as outlined in the guidelines provided by Gagarina et al. (2012) for the English narratives and Blom and De Jong (2012) for the Dutch narratives. To explain the task, a blank MAIN story strip was used. It was explained to the children that they were expected to tell a story based on the six pictures they would see on the strip of paper they would receive. The child was allowed to scrutinise all six pictures before telling the story and was not allowed to show the experimenter the pictures. Most of the Dutch stories were told without many questions or concerns. The English stories were more challenging for the children. Most comments during their stories concerned not knowing words in English. In these cases, the participants were instructed that they could use a word they thought was the right one, make one up or use the Dutch word for it.

The stories were recorded and later transcribed by four trained university students in the CHAT (Codes for the Human Analysis of Transcripts) transcription system using CLAN (Computerised Language Analysis) software (MacWhinney, 2000). These students were Dutch native speakers studying towards a degree in English linguistics. Reliability was checked by producing an independent transcription accuracy. All transcriptions were checked by another trained research assistant, and an agreement score was calculated by dividing the total number of tokens by the total number of agreements which resulted in a score of 92.7%. All inconsistencies and cases of disagreement were re-evaluated and, if needed, corrected.

Scoring. The transcribed utterances were transformed to C-units following the guidelines provided by Curenton (2004). An additional guideline that was adhered to specifically for the English narratives concerned C-units containing code-switches. If a C-unit

[2] *Gisteren* is Dutch for *yesterday*.

contained only code-switches, it was removed from the transcript. In all other cases, the C-unit was kept as part of the transcript in order to give credit to all the English produced by the children.

Each C-unit was scrutinised for the presence of grammatical errors. The following types of errors were attested in both Dutch and English narratives: verb errors (omission and substitution), determiner errors (omission, substitution), preposition errors, pronoun errors (gender, number), null subject sentences and word order errors. In addition, the Dutch narratives also contained gender errors. To control for differences in talkativeness, error frequencies were normalised per number of tokens in the target language. In the case of word order errors, normalisation involved obtaining error rates per C-unit.

3.6. Data analysis

The GJT data were analysed by means of generalised mixed-effects logistic regression analyses using the *glmer* function in R (Bates et al., 2015; R Development Core Team, 2008). The performance of the CG at pre-test was taken as the baseline. Participant nested in Class (Class:Participant) and Item were included in the random part of all the models. Narrative data was analysed by means of multilevel linear regression analyses performed using the lmerTest package in R (Kuznetsova et al., 2017). Story (Cat; Dog) and Participant nested in Class (Class:Participant) were included in the random part of the model. In all the analyses, all fixed and random effects were entered at once because the reported models are the minimal models that are required to answer the corresponding research questions.

We computed several measures of effect size that are considered appropriate for multilevel models (Lorah, 2018). Nakagawa's marginal and conditional R^2 were calculated using the MuMIn package in R. The marginal R^2 captures the variance of the fixed effects, while the conditional R^2 reflects the variance explained by the entire model (Nakagawa & Schielzeth, 2013). Based on these two values, Cohen's f^2 – an effect size measure related to variance explained – was computed using the following formula: $f^2 = \frac{R^2}{1-R^2}$. A rule of thumb for the interpretation of f^2 is as follows: 0.02 is a small effect, 0.15 is a medium effect and 0.35 is a large effect (Cohen, 1992; Lorah, 2018). Finally, to determine the variance explained by the individual fixed predictors, R^2_β and the corresponding confidence intervals for each fixed effect were obtained using the Kenward-Roger approach in the r2glmm package in R.

4. Results

4.1. Grammaticality Judgment Task

Since grammatical and ungrammatical sentences have been shown to measure different constructs in L2 learners (Ellis, 2005), these two types of sentences were analysed separately. Recall that the participants could easily score a point for judging grammatical sentences correct, but in order to be awarded a point on ungrammatical trials, they also had to identify or correct the error. So, the ungrammatical trials are more informative about grammatical competence.

The performance of the two groups over time is presented in Fig. 2. The model coefficients are reported in Table 2 for grammatical sentences and in Table 3 for ungrammatical sentences. The performance of both groups on the grammatical sentences was very high. As shown in Table 2, there were no differences between groups at pre-test (parameter 1). The performance of the CG improved with time (parameter 2), which was also the case in the IG, as evidenced by the non-significant interaction (parameter 3). Hence, both the

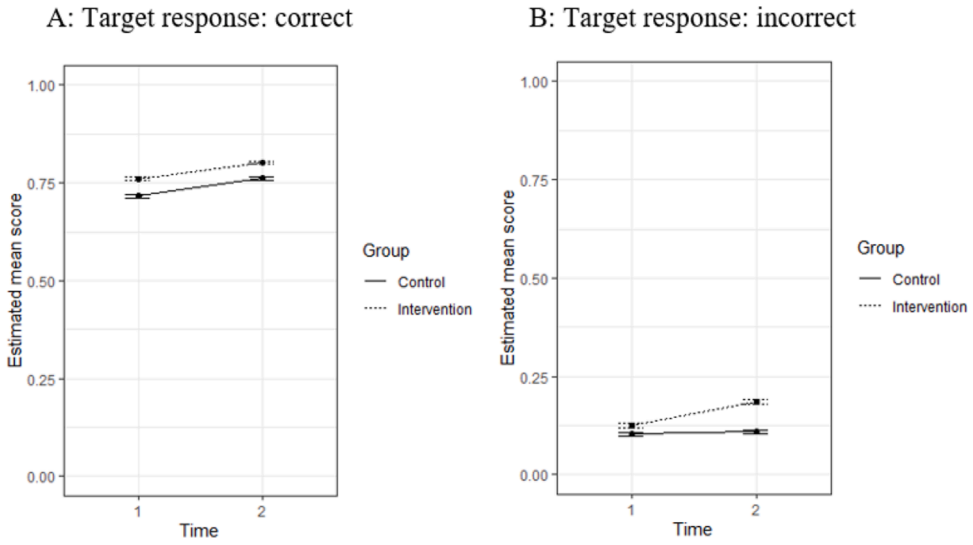


Fig. 2. Mean proportion of correct responses on grammatical sentences (panel A) and ungrammatical sentences (panel B), by group and time (the error bars represent SE 95% CI).

intervention and the business-as-usual EFL lessons resulted in improvements in the ability to detect grammatically correct sentences in L2 English. The entire model explained 19% of variance (conditional $R^2 = 0.19$). The marginal R^2 is 0.01, and f^2 for the fixed effects is 0.01 (negligible ES).

The results for ungrammatical sentences were markedly different. Both groups scored low at pre-test. The model coefficients in Table 3 indicate that there were no differences between groups at Time 1 (parameter 1). The performance of the CG did not change with time (parameter 2). In contrast, the performance of the IG improved from pre- to post-test (parameter 3), which is also confirmed by an additional analysis with the IG as the baseline ($B = 0.67$, $SE = 0.14$, $z = 4.63$, $p < .001$). The conditional R^2 for the entire model is 0.506, i.e., the model explains 51% of variance. The marginal R^2 is 0.015, and f^2 for the fixed part of the model is 0.02 (small ES).

Fig. 3 presents the pre- to post-test gain scores of individual children in the two groups. It is evident from the figure that there were more children with higher gains in the IG than in the CG. At the same time, there were four children in the CG who made remarkable progress without explicit instruction. There was one clear outlier in the IG, a monolingual Dutch-speaking child with lower-than-average PPVT (71) and CELF (55) scores. Overall, there appears to be a divide in the IG between pupils who respond well and those not responding to the intervention. We will revisit this observation in the Discussion.

To determine whether the IG only improved on the grammatical categories that were treated in the intervention (as opposed to filler categories that were not targeted in the intervention), we created a model predicting accurate performance on ungrammatical sentences from Category (Treated; Fillers), Group (Intervention; Control), Time (T1; T2) and the interaction between the three main effects. The model comparisons are presented in Table 4. At Time 1 the CG did not perform better on the categories selected for the intervention than on fillers (parameter 1). The lack of significant interaction between Group and Category (parameter 4) demonstrates that that was also the case for the IG at Time 1. The IG at Time 1 did not perform better than the CG on the items selected for the intervention (parameter 2). There was no improvement in the CG on the grammatical categories treated in the intervention (parameter 3) and this was also the case for the filler categories (parameter 5). The significant interaction between Group and Time2 (parameter 6) shows that the IG made more progress on the treated categories. But the lack of improvement on the non-treated/filler categories (as attested for the CG) also holds for the IG, as evidenced by the non-significant three-way interaction in parameter 7. The entire model explained 51% of variance (conditional $R^2 = 0.506$). The marginal R^2 is 0.043, and f^2 for the fixed part of the model is 0.04 (small ES).

In the final analysis we focused on the IG only and compared their performance on familiar sentences practised in the lessons and new sentences from the treated grammatical categories (ungrammatical trials only). As shown in Table 5, at pre-test there were no differences between the two sentence types (parameter 1). The performance on the familiar sentences significantly improved (parameter 2), and the progress was equally strong for unfamiliar sentences (parameter 3). The entire model explained 62% of variance (conditional $R^2 = 0.618$). The marginal R^2 is 0.006, and f^2 for the fixed part of the model is 0.01 (negligible ES).

In summary, there were no differences between groups at pre-test. The performance of both groups improved on grammatical sentences, but only the IG improved on ungrammatical sentences. However, the effects were small. The pre- to post-test progress in the IG was significant for grammatical categories treated in the intervention but not for untreated/filler categories. Within the treated categories, the progress was equally strong for familiar sentences that had been practised in the lessons and unfamiliar sentences.

4.2. Grammatical accuracy in the narratives

First, we predicted the number of grammatical errors in the narratives from the GJT scores. The GJT scores did not predict the accuracy of Dutch narratives ($B = -7.79$, $SE = 7.44$, $t = -1.05$, $p < .298$; conditional $R^2 = 0.501$, marginal $R^2 = 0.009$, $f^2 = 0.01$). But they did predict the number of errors in the English narratives: The higher the GJT scores were, the fewer errors the narratives contained ($B = -0.007$, $SE = 0.002$, $t = -3.16$, $p = .002$, conditional $R^2 = 0.340$, marginal $R^2 = 0.091$, $f^2 = 0.10$).

Then we compared the pre- to post-test change in the total error rates in the two groups by creating models with Group (Intervention; Control), Time (T1; T2) and their interaction as main effects. The performance of the CG at Time 1 was taken as the baseline. The results for Dutch and English are summarised in Tables 6 and 7, respectively. Detailed descriptive statistics for each error category can be found in the Supplementary Materials.

Regarding error rates in Dutch, the groups did not differ at pre-test (parameter 1 in Table 6). In the CG, there was no decrease in error frequencies with time (parameter 2), whereas, in the IG, there was (parameter 3), which is also supported by an additional analysis with the IG as the baseline ($B = -0.03$, $SE = 0.01$, $t = -3.25$, $p = .002$). The entire model explained 53% of variance (conditional $R^2 = 0.528$). The marginal R^2 is 0.044, and f^2 for the fixed part of the model is 0.05 (small ES). The R^2_β values in Table 6 show that the strongest predictor in the model was the interaction between Group and Time. The advantage of the IG is also evident in the individual gain scores in Fig. 4. The mean gain score was higher in the IG, and the proportion of children with higher gain scores was also higher in the IG. The analysis of specific error categories revealed that none of them showed improvement in the CG. The performance of the IG significantly improved only in the use of verb inflection ($B = -0.01$, $SE = 0.002$, $t = -2.45$, $p = .031$; conditional $R^2 = 0.708$, marginal

Table 2
Coefficients of the comparisons between groups for grammatical sentences.

	B	*SE*	z	*p*	R^2_β [95% *CI*]
(Intercept)	1.06	0.17	6.08	< .001	
1. GroupIntervention	0.29	0.18	1.60	.110	.001 [.000, .005]
2. Time2	0.25	0.11	2.25	.025	.001 [.001, .004]
3. GroupIntervention*Time2	0.02	1.16	0.10	.917	.000 [.000, .001]

Table 3
Coefficients of the comparisons between groups for ungrammatical sentences.

	B	*SE*	z	*p*	R^2_β [95% *CI*]
(Intercept)	-3.09	0.35	-8.74	<.001	
1. GroupIntervention	0.12	0.40	0.29	.770	.000 [.000, .001]
2. Time2	0.09	0.17	0.52	.601	.000 [.001, .004]
3. GroupIntervention*Time2	0.58	0.22	2.58	.010	.001 [.000, .004]

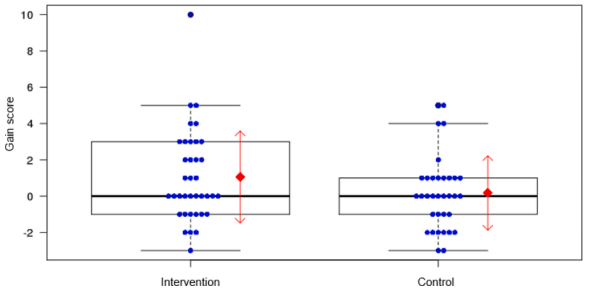


Fig. 3. The distribution of gain scores on the ungrammatical trials of the GJT (means and +/-1 SDs are displayed in red).

Table 4
Model coefficients for comparisons between groups on treated and non-treated categories, over time (ungrammatical items only).

	B	*SE*	z	*p*	R^2_β [95% *CI*]
(Intercept)	-2.85	0.37	-7.69	<.001	
1. CategoryFillers	-0.63	0.45	-1.40	.162	.001 [.000, .004]
2. GroupIntervention	0.21	0.41	0.51	.613	.000 [.000, .002]
3. Time2	0.16	0.19	0.82	.414	.000 [.000, .002]
4. CategoryFillers*GroupIntervention	-0.40	0.37	-1.07	.285	.000 [.000, .002]
5. CategoryFillers*Time2	-0.28	0.39	-0.72	.472	.000 [.000, .002]
6. GroupIntervention*Time2	0.51	0.25	2.02	.044	.000 [.000, .003]
7. CategoryFillers*GroupIntervention*Time2	0.29	0.51	0.57	.567	.000 [.000, .001]

Table 5
Coefficients of the comparisons between familiar and new sentences.

	B	*SE*	z	*p*	R^2_β [95% *CI*]
(Intercept)	-0.46	0.63	-0.73	.466	
1. TypeUnfamiliar	0.21	0.74	0.28	.781	.000 [.000, .003]
2. Time2	0.62	0.19	3.26	.001	.002 [.000, .007]
3. TypeUnfamiliar*Time2	-0.27	0.23	-1.20	.229	.000 [.000, .003]

Table 6
Coefficients of the comparisons between groups for normalised error frequencies in the Dutch narratives over time.

	B	*SE*	t	*p*	R^2_β [95% *CI*]
(Intercept)	0.08	0.01	8.62	< .001	
1. GroupIntervention	0.02	0.01	1.97	.051	.011 [.000, .108]
2. Time2	0.002	0.01	0.28	.784	.058 [.001, .201]
3. GroupIntervention*Time2	-0.03	0.01	-2.45	.017	.081 [.004, .234]

Table 7
Coefficients of the comparisons between groups for error rates in the English narratives over time.

	B	*SE*	*t*	*p* value	R^2_β [95% *CI*]
(Intercept)	0.20	0.03	5.98	<.001	
1. GroupIntervention	0.05	0.05	1.27	.207	.014 [.000, .089]
2. Time2	-0.05	0.04	-1.38	.172	.013 [.000, .085]
3. GroupIntervention*Time2	-0.04	0.05	-0.76	.453	.004 [.000, .060]

$R^2 = 0.021$, $f^2 = 0.02$).

Regarding error rates in English, Fig. 5 shows that the gains were on average higher in the IG. However, at the group level, this progress was not strong enough to result in a significant Group by Time interaction (Table 7). It is also noteworthy that, within the IG, there were two clusters – children who seemed to benefit from the intervention and those who did not.

5. Discussion

This paper reports a nonrandomised controlled trial testing the efficacy of a new teaching method tailored to the specific needs of foreign language learners with DLD. The key features of our remedial approach include metalinguistic instruction, explicit cross-linguistic comparisons and multimodal interaction with the material. The results demonstrate that the CodeTaal approach had a positive effect on error detection in the IG, as measured by the GJT. Whereas both groups showed significant progress on grammatical sentences, only the IG significantly improved in their ability to identify ungrammaticalities in English, but the effect was small. Lack of differences between familiar and unfamiliar sentences revealed that the intervention participants were able to generalise the rules to new instances of the treated grammatical categories, which is in line with prior intervention studies using explicit multimodal approaches in L1 therapy for children with DLD (Bolderson et al., 2011; Calder et al., 2021; Ebbels et al., 2007; Finestack, 2018; Levy & Friedmann, 2009; Seeff-Gabriel et al., 2012).

The effects of the intervention were also evident in the fact that enhanced performance was observed for the grammatical categories treated in the intervention (Present Simple, determiners, plurals, SVO) but not for the untreated categories (Present Progressive, Past Simple, Present Perfect). This result is consonant with prior research showing that metalinguistic interventions lead to enhanced performance on the directly targeted categories but not to improvement in non-treated categories (Calder et al., 2021; Ebbels et al., 2007). We had planned to continue with the treatment of these initially untreated (filler) categories during the second stage of the intervention in the spring term. However, after nine lessons, covering Present Continuous and Past Simple, Stage 2 of the intervention was discontinued due to the pandemic, and the planned post-tests could not be administered. We could only administer the GJT again after the summer break (September 2020) and only to a subset of the intervention participants (n=24). The results revealed a significant progress on the original filler categories (i.e., the categories treated in Stage 2 of the intervention) between the second and the third measurement ($B = -.51$, $SE = 0.20$, $z = 2.55$, $p = .011$). At the same time, there was no improvement on the categories treated in Stage 1 ($B = 0.14$, $SE = 0.14$, $z = 0.99$, $p = .320$). Even though this third measurement should be taken with caution due to several confounds, the findings are still indicative when considered in combination with the current results. The pupils appear to respond to the grammatical structures targeted in the intervention, but growth may stagnate once new topics are introduced into the lessons. Importantly, the gains obtained in Stage 1, albeit small, were maintained, even after the lockdown and the summer break.

The advantage of the IG in the ability to identify ungrammaticality did not translate into higher accuracy rates in semi-spontaneous production in English. Even though there was a subgroup of the IG that made good progress, no advantage of the intervention could be posited at the group level. These results run counter to the findings reported by Calder et al. (2021). In their study, the IG showed improvement on the elicited production of past tense but not on a GJT. The different results might be due to the test instruments: Calder et al. used an elicitation task specifically targeting verb morphology. In contrast, a narrative task, as used in our study, is more taxing since it involves not only language production at the microlevel but also constructing a coherent discourse at the macrolevel. Such complex language tasks may take longer than 12 lessons to show learning gains from the intervention. The different pattern for the GJT results may also be due to the fact that our participants were 9–13 years old, whereas Calder et al.'s participants were younger (age 6–7). It is plausible that older children have an advantage in the development of metalinguistic judgments due to their more developed cognitive skills and ameliorating declarative memory in particular. Finally, the participants in Calder et al.'s study were native speakers acquiring English in a naturalistic setting, whereas our participants were learning English as a FL in a school setting with limited exposure. In naturalistic L1/L2 acquisition, implicit grammar knowledge usually develops before metalinguistic awareness. In contrast, EFL learning in instructed settings with minimal classroom exposure more often involves the development of explicit grammar knowledge that may become proceduralised later, through practice (DeKeyser, 2007). Hence, explicit interventions in naturalistic settings are likely to boost the development of implicit grammar knowledge in the first place, whereas a similar approach in an EFL setting primarily leads to enhancement of metalinguistic knowledge, which might facilitate grammar in use at a later stage. This possibility is further reinforced by our finding that the GJT scores predicted accuracy of the English narratives. This

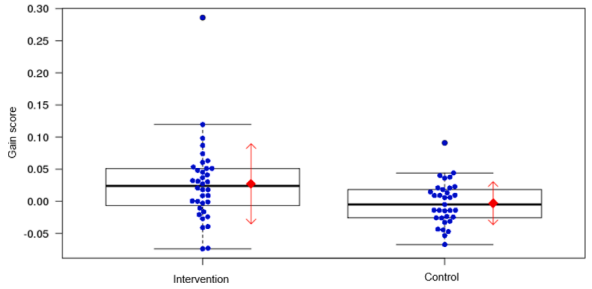


Fig. 4. The distribution of gain scores for normalised error frequencies in the Dutch narratives (means and +/-1 SDs are displayed in red).

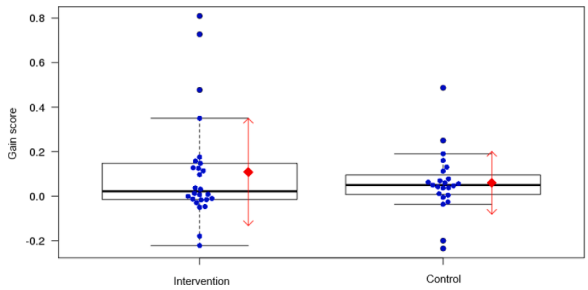


Fig. 5. The distribution of gain scores for normalised error frequencies in the English narratives (means and +/-1 SDs are displayed in red).

scenario would be in line with the results of a two-year intervention study reported in Piggott (2019), in which the accuracy advantage of an explicitly taught group only showed after two years of explicit grammar lessons. Hence, even though based on our data we cannot claim that our intervention had a positive effect on the use of grammar in language production, the fact that it did enhance the ability to detect errors and that this ability was predictive of accuracy rates makes it plausible that the advantages in the narrative task would have become visible if the intervention had lasted longer, or if we had administered a delayed post-test as originally planned. The finding that performance in the GJT predicts accuracy in production supports earlier research with L1 speakers of English (Rice et al., 1999; Smith-Lock, 1995) and extends these findings to the (E)FL setting.

Our findings also converge with research demonstrating that instructional approaches including explicit instruction about L1 enhance metalinguistic awareness and lead to greater learning gains in an L2 (Lucas, 2020; McManus, 2019; McManus & Marsden, 2017). Interestingly and importantly, our intervention had a positive effect on grammatical accuracy in Dutch. The error rates in the CG did not change over time. In contrast, the narratives of the IG at post-test were more accurate than their stories elicited at pre-test. The category that showed most improvement was verb inflection, which is a clinical marker of DLD in Dutch (Blom et al., 2013). In each lesson the pupils were coding sentences in CodeTaal, where the morpheme shapes had to be attached to the verb stems in both languages. This multimodal metalinguistic instruction on Dutch grammar resulted in improved use of verbal morphology in the majority language, which is in line with prior studies demonstrating the effectiveness of metalinguistic instruction and visual coding for enhancing verb use in the L1 (Calder et al., 2018, 2021; Ebbels, 2007; Tobin & Ebbels, 2019). Our findings reveal that such improvement is possible even if children are taught in groups and even if a foreign language is the focus of the lesson. At the group level, the effect was small. But the analysis of individual gain scores revealed that there were more children with larger progress in the IG than in the CG.

Based on the current results, we cannot separate the effects of explicit rule instruction and the effects of cross-linguistic comparisons. Future research could tease the two strategies apart to determine their individual contribution. Likewise, based on our results, it is not clear whether similar gains could be achieved without visual and tactile support. There is evidence in the literature that metalinguistic instruction alone may be sufficient (Finestack, 2018; Hirschman, 2000). Therefore, it is worth investigating whether metalinguistic contrastive instruction alone will suffice for facilitating FL learning by children with DLD in a school setting. It is possible that visual support is important in primary school but can be abandoned when teaching older (secondary-school) pupils with DLD.

A significant limitation of this study is that randomisation was not possible. This inevitably affected internal validity and generalisability of the results, since differences in progress could also be due to school differences. Controlling for school and class differences in the random part of the models, as well as the fact that the schools were matched for a number of crucial characteristics (amount and length of EFL instruction; similar school philosophy and demographics) mitigates the risks and suggests that the advantages in the IG are likely to be due to the intervention. Even though randomisation is more challenging in classroom settings compared to interventions delivered in one-to-one sessions, future studies should use more robust randomised control trials in tandem with blinding procedures to corroborate the conclusions of this early efficacy study.

Another important direction to pursue would be to investigate whether certain types of learners are more receptive to approaches such as CodeTaal. The scrutiny of individual gains showed that some of the intervention participants seemed to respond very well and made significant progress, whereas others did not improve their English or Dutch skills. Prior research on L1 therapy has repeatedly shown that there are significant individual differences in responsiveness to metalinguistic interventions (Ebbels, 2007; Ebbels & Van der Lely, 2001; Guendouzi, 2003; Hirschman, 2000; Spooner, 2002). For example, it is possible that the CodeTaal approach would work best for learners who have a stronger declarative memory because explicit language learning is largely supported by the declarative memory system. At the same time, children with weaker procedural memory may achieve greater gains with this method because it will help them to circumvent their procedural learning disadvantage by learning through declarative memory (cf. Hirschman, 2000). It is also possible that learners who are more proficient in the school language will benefit more because our remedial approach capitalises on positive transfer from the language of instruction (cf. Armon-Lotem et al., 2021).

This study only partly fits the definition of a Tier 1 intervention study (Ebbels et al., 2019) because it only included pupils with language disorders (rather than mixed-ability classes). This study provided an indication that CodeTaal might be a promising teaching approach for special education. It remains to be seen whether it can also be meaningfully implemented in a regular educational setting where children with and without DLD share classrooms.

6. Conclusion

The results of this early efficacy study demonstrate that, if provided with adequate support, pupils with DLD are able to make progress in foreign language learning in a classroom setting. At the group level, twelve 45-min lessons taught to groups of 13–15 pupils were sufficient to make significant progress in detecting ungrammaticalities in English, and also had a positive effect on the production of verb morphology in the majority language. However, the effects were small and there was significant variability in responsiveness to the intervention.

CRediT authorship contribution statement

Elena Tribushinina: Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Project administration, Supervision, Validation, Visualization, Writing – original draft, Writing – review & editing. **Geke Niemann:** Conceptualization, Investigation, Methodology, Validation, Writing – review & editing. **Joyce Meuwissen:** Conceptualization, Funding acquisition, Investigation, Methodology, Supervision, Writing – review & editing. **Megan Mackaaij:** Data curation, Project administration, Visualization, Writing – original draft, Writing – review & editing. **Gabriëlla Lahdo:** Investigation, Methodology, Writing – review & editing.

Acknowledgments

We are very grateful to the two anonymous reviewers for their valuable constructive feedback. We also would like to thank the participating schools, parents and children that made this investigation possible. We thank Angelique Niemann-Jager for creating the intervention materials and the test materials. Desiree Mastellone, Esther Paasman and Anna Drukker kindly helped with data collection.

Funding

This work was supported by the Netherlands Organization for Scientific Research (NWO) under Grant 015.015.061 to the first author.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.jcomdis.2022.106269.

References

Ammar, A., Lightbown, P. M., & Spada, N. (2010). Awareness of L1/L2 differences: Does it matter? *Language Awareness, 19*(2), 129–146.
Armon-Lotem, S., Restrepo, M. A., Lipner, M., Ahituv-Shlomo, P., & Altman, C. (2021). Vocabulary gains in bilingual narrative intervention. *Language, Speech, and Hearing Services in Schools, 52*(1), 436–448.
Balthazar, C. H., Ebbels, S., & Zwitserlood, R. (2020). Explicit grammatical intervention for developmental language disorder: Three approaches. *Language, Speech, and Hearing Services in Schools, 51*(2), 226–246.
Balthazar, C. H., & Scott, C. M. (2018). Targeting complex sentences in older school children with specific language impairment: Results from an early phase treatment study. *Journal of Speech, Language, and Hearing Research, 61*(3), 713–728.
Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software, 67*, 1–48.
Bell, P., Fortier, V., & Gauvin, I. (2020). Using L1 knowledge about language during L2 error correction: Do students make cross-linguistic connections? *Language Awareness, 29*(2), 95–113.
Benson, S., & DeKeyser, R. (2019). Effects of written corrective feedback and language aptitude on verb tense accuracy. *Language Teaching Research, 23*(6), 702–726.
Blom, W., & De Jong, J. D (2012). MAIN: Nederlandse versie (Meertalig Assessment Instrument voor Narratieven). *Paper in Linguistics*, 2–41.
Blom, E., De Jong, J., Orgassa, A., Baker, A., & Weerman, F. (2013). Verb inflection in monolingual Dutch and sequential bilingual Turkish-Dutch children with and without SLI. *International Journal of Language and Communications Disorders, 48*(4), 382–393.
Blom, E., & Paradis, J. (2015). Sources of individual differences in the acquisition of tense inflection by English second language learners with and without specific language impairment. *Applied Psycholinguistics, 36*(4), 953–976. https://doi.org/10.1017/S014271641300057X
Bolderson, S., Dosanjh, C., Milligan, C., Pring, T., & Chiat, S. (2011). Colourful semantics: A clinical investigation. *Child Language Teaching and Therapy, 27*(3), 344–353.
Bouffard, L. A., & Sarkar, M. (2008). Training 8-year-old French immersion students in metalinguistic analysis: An innovation in form-focussed pedagogy. *Language Awareness, 17*(1), 3–24.
Bryan, A. (1997). Colourful semantics. In S. Chiat, J. Law, & J. Marshall (Eds.), *Language disorders in children and adults: Psycholinguistic approaches to therapy* (pp. 143–161). Whurr.
Calder, S. D., Claessen, M., & Leitão, S. (2018). Combining implicit and explicit intervention approaches to target grammar in young children with Developmental Language Disorder. *Child Language Teaching and Therapy, 34*(2), 171–189.
Calder, S. D., Claessen, M., Ebbels, S., & Leitão, S. (2020). Explicit grammar intervention in young school-aged children with Developmental Language Disorder: An efficacy study using single-case experimental design. *Language, Speech, and Hearing Services in Schools, 51*(2), 296–316.
Calder, S. D., Claessen, M., Ebbels, S., & Leitão, S. (2021). The efficacy of an explicit intervention approach to improve past tense marking for early school-age children with developmental language disorder. *Journal of Speech, Language, and Hearing Research, 64*(1), 91–104.
Cohen, J. (1992). A power primer. *Psychological Bulletin, 112*, 155–159.
Curenton, S. M. (2004). The association between narratives and theory of mind for low- income pre-schoolers. *Early Education and Development, 15*(2), 124–146.

Dam, Q., Pham, G. T., Pruitt-Lord, S., Limon-Hernandez, J., & Goodwiler, C. (2020). Capitalizing on cross-language similarities in intervention with bilingual children. *Journal of Communication Disorders, 87*, Article 106004.
DeKeyser, R. (2007). Skill acquisition theory. In B. VanPatten, & J. Williams (Eds.), *Theories in second language acquisition* (pp. 97–113). Erlbaum.
Donovan, M. S., & Bransford, J. D. (2005). *How students learn: History, mathematics, and science in the classroom*. National Academic Press.
Ebbels, S. H. (2007). Teaching grammar to school-aged children with Specific Language Impairment using Shape Coding. *Child Language Teaching and Therapy, 23*(1), 67–93.
Ebbels, S. H. (2014). Effectiveness of intervention for grammar in school-aged children with primary language impairments: A review of the evidence. *Child Language Teaching and Therapy, 30*(1), 7–40.
Ebbels, S. H., McCartney, E., Slonims, V., Dockrell, J. E., & Norbury, C. F. (2019). Evidence-based pathways to intervention for children with language disorders. *International Journal of Language and Communication Disorders, 54*(1), 3–19. https://doi.org/10.1111/1460-6984.12387
Ebbels, S. H., & Van der Lely, H. K. (2001). Meta-syntactic therapy using visual coding for children with severe persistent SLI. *International Journal of Language and Communication Disorders, 36*, 345–350.
Ebbels, S. H., Van der Lely, H., & Dockrell, J. E. (2007). Intervention for verb argument structure in children with persistent SLI: A randomized control trial. *Journal of Speech, Language, and Hearing Research, 50*, 1330–1349.
Ellis, R. (2005). Measuring implicit and explicit knowledge of a second language: A psychometric study. *Studies in Second Language Acquisition, 27*, 141–172.
Ellis Weismer, S., Evans, J., & Hesketh, L (1999). An examination of verbal working memory capacity in children with SLI. *Journal of Speech, Language, and Hearing Research, 42*, 1249–1260.
Erlam, R. (2005). Language aptitude and its relationship to instructional effectiveness in second language acquisition. *Language Teaching Research, 9*(2), 147–171.
Eurostat. (2019). *Foreign language learning statistics*. Retrieved on 23 February 2022 from http://ec.europa.eu/eurostat/statistics-explained/index.php/Foreign_language_learning_statistics.
Evans, J., Saffran, J. R., & Robe-Torres, K. (2009). Statistical learning in children with specific language impairment. *Journal of Speech, Language, and Hearing Research, 52*, 321–335.
Finestack, L. H. (2018). Evaluation of an explicit intervention to teach novel grammatical forms to children with developmental language disorder. *Journal of Speech, Language, and Hearing Research, 61*(8), 2062–2075.
Frizelle, P., Tolonen, A.-K., Tulip, J., Murphy, C. A., Saldana, D., & McKean, C. (2021). The impact of intervention dose form on oral language outcomes for children with developmental language disorder. *Journal of Speech, Language, and Hearing Research, 64*, 3253–3288.
Fumero, K., & Tibi, S. (2020). The importance of morphological awareness in bilingual language and literacy skills: Clinical implications for speech-language pathologists. *Language, Speech, and Hearing Services in Schools, 51*(3), 572–588.
Gagarina, N., Klop, D., Kunnari, S., Tantele, K., Valimaa, T., Balciuniene, I., Bohnacker, U., & Walters, J. (2012). MAIN: Multilingual assessment instrument for narratives. *ZAS Papers in Linguistics, 56*. Retrieved from http://www.zas.gwz-berlin.de/zaspil56.html.
Garton, S., Copland, F., & Burns, A. (2011). Investigating global practices in teaching English to young learners. *ELT Research Papers, 11*(1), 1–24.
Gopnik, M., & Crago, M. (1991). Familial aggregation of a developmental language disorder. *Cognition, 39*, 1–50.
Gottardo, A., Stanovich, K. E., & Siegel, L. S. (1996). The relationships between phonological sensitivity, syntactic processing, and verbal working memory in the reading performance of third-grade children. *Journal of Experimental Child Psychology, 63*(3), 563–582. https://doi.org/10.1006/jecp.1996.0062
Guendouzi, J. (2003). SLI', a generic category of language impairment that emerges from specific differences: A case study of two individual linguistic profiles. *Clinical Linguistics & Phonetics, 17*(2), 135–152.
Gursoy, E., Korkmaz, S. C., & Damar, E. A. (2017). English language teaching within the new educational policy of Turkey: Views of stakeholders. *International Education Studies, 10*(4), 18–30.
Haamberg, R., Hofman, K., Maaswinkel, E., & Rödiger, A. (2008). Doeltaal-voertaal: Een handreiking voor mtv-docenten. *Levende Talen Magazine, 95*(1), 5–8.
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. https://etheses.whiterose.ac.uk/.
Hirschman, M. (2000). Language repair via metalinguistic means. *International Journal of Language and Communication Disorders, 35*(2), 251–268.
Hulstijn, J. (2005). Theoretical and empirical issues in the study of implicit and explicit second-language learning. *Studies in Second Language Acquisition, 27*, 129–140.
Hwu, F., Pan, W., & Sun, S. (2014). Aptitude-treatment interaction effects on explicit rule learning: A latent growth curve analysis. *Language Teaching Research, 18*(3), 294–319.
Kamhi, A. G., & Koenig, L. A. (1985). Metalinguistic awareness in normal and language-disordered children. *Language, Speech and Hearing Services in Schools, 16*(3), 199–210.
Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. (2017). Lmer package: Tests in linear mixed effects models. *Journal of Statistical Software, 82*(13), 1–26.
Lammertink, I., Boersma, P., Wijnen, F., & Rispens, J. (2017). Statistical learning in specific language impairment: A meta-analysis. *Journal of Speech, Language, and Hearing Research, 60*(12), 3474–3486.
Lammertink, I., Boersma, P., Wijnen, F., & Rispens, J. (2020). Statistical learning in the visuomotor domain and its relation to grammatical proficiency in children with and without developmental language disorder: A conceptual replication and meta-analysis. *Language Learning and Development, 16*(4), 426–450.
Leonard, L. B. (2014). *Children with Specific Language Impairment*. MIT Press.
Levy, H., & Friedmann, N. (2009). Treatment of syntactic movement in syntactic SLI: A case study. *First Language, 29*(1), 15–50.
Lichtman, K. (2016). Age and learning environment: Are children implicit second language learners? *Journal of Child Language, 43*(3), 707–730.
Lightbown, P., & Spada, N. (2000). Do they know what they are doing? L2 learners' awareness of L1influence. *Language Awareness, 9*(4), 198–217.
Littlewood, W., & Yu, B. (2011). First language and target language in the foreign language classroom. *Language Teaching, 44*, 64–77.
Lorah, J. (2018). Effect size measures for multilevel models: Definition, interpretation, and TIMSS example. *Large-scale Assessments in Education, 6*(8), 1–11.
Lucas, M. (2020). Cross-linguistic influence and plural marking: Enhancing error recognition and written accuracy among Japanese EFL learners through web-based awareness-raising. *Language Awareness, 29*(2), 155–177.
Lukács, Á., Kemény, F., Lum, J. A. G., & Ullman, M. T. (2017). Learning and overnight reten-tion in declarative memory in specific language impairment. *PLoS ONE*, (1), 12. https://doi.org/10.1371/journal.pone.0169474
Lum, J. A., Conti-Ramsden, G., Page, D., & Ullman, M. (2012). Working, declarative and procedural memory in specific language impairment. *Cortex, 48*(9), 1138–1154.
Lum, J. A., Conti-Ramsden, G., Morgan, A. T., & Ullman, M. (2014). Procedural learning deficits in specific language impairment (SLI): A meta-analysis of serial reaction time task performance. *Cortex, 51*, 1–10.
MacWhinney, B. (2000). *The CHILDES Project: Tools for Analyzing Talk* (3rd Edition). Lawrence Erlbaum Associates.
McDonald, J. L. (2008). Grammaticality judgments in children: The role of age, working memory and phonological ability. *Journal of Child Language, 35*(2), 247–268.
McManus, K. (2019). Awareness of L1 form-meaning mappings can reduce crosslinguistic effects in L2 grammatical learning. *Language Awareness, 28*(2), 114–138.
McManus, K. (2022). *Crosslinguistic influence and second language learning*. Routledge.
McManus, K., & Marsden, E. (2017). L1 explicit instruction can improve L2 online and offline performance. *Studies in Second Language Acquisition, 39*(3), 459–492.
McManus, K., & Marsden, E. (2018). Online and offline effects of L1 practice in L2 grammar learning: A partial replication. *Studies in Second Language Acquisition, 2*, 459–475.
McManus, K., & Marsden, E. (2019). Using explicit instruction about L1 to reduct crosslinguistic effects in L2 grammar learning: Evidence from oral production in L2 French. *The Modern Language Journal, 103*(2), 459–480.
Montessori, M. (1912). *The Montessori method: Scientific pedagogy as applied to child education in 'the children's houses' with additions and revisions by the author (AE George, Trans.)*. Robert Bentley.
Nakagawa, S., & Schielzeth, H. (2013). A general and simple method for obtaining R^2 from generalized linear mixed-effects models. *Methods in Ecology and Evolution, 4*, 133–142.

Norris, J. M., & Ortega, L. (2000). Effectiveness of L2 instruction: A research synthesis and quantitative meta-analysis. *Language Learning, 50*(3), 417–528.
Obeid, R., Brooks, P., Powers, K., Gillespie-Lynch, K., & Lum, J. A. (2016). Statistical learning in Specific Language Impairment and Autism Spectrum Disorder: A meta-analysis. *Frontiers in Psychology, 7*, 1–18.
Paradis, J., Genesee, F., & Crago, M. (2021). *Dual language development and disorders: A handbook on bilingualism and second language learning*. Brookes.
Perozzi, J. A., & Sanchez, M. L. C. (1992). The effect of instruction in L1 on receptive acquisition of L2 for bilingual children with language delay. *Language, Speech, and Hearing Services in Schools, 23*(4), 348–352.
Piggott, L. (2019). *First meaning then form: A longitudinal study on the effects of delaying and reducing explicit form-focused instruction for young adolescent EFL learners*. Utrecht: LOT Dissertation Series.
Rao, Z., & Yu, P. (2019). Teaching English as a foreign language to primary school students in East Asia: Challenges and future prospects. *English Today, 35*(3), 16–21.
R Development Core Team. (2008). *R: A language and environment for statistical computing [Computer software manual]*. http://www.R-project.org.
Rebuschat, P., & Williams, J. N. (2012). Implicit and explicit knowledge in second language acquisition. *Applied Psycholinguistics, 33*, 829–856.
Rice, M. L. (2000). Grammatical symptoms of specific language impairment. In D. V. M. Bishop, & L. B. Leonard (Eds.), *Speech and language impairments in children: Causes, characteristics, intervention and outcome* (pp. 17–34). Psychology Press.
Rice, M. L., Wexler, K., & Redmond, S. M. (1999). Grammaticality judgements of an extended optional infinitive grammar: Evidence from English-speaking children with specific language impairment. *Journal of Speech Language and Hearing Research, 42*(4), 943–961. https://doi.org/10.1044/jslhr.4204.943
Rose, P. (2016). *Engels in het basisonderwijs: Domeinbeschrijving ten behoeve van peilingsonderzoek*. SLO. Retrieved on 19 June 2022 from https://www.slo.nl/publicaties/@4559/engels-1/.
Seeff-Gabriel, B., Chiat, S., & Pring, T. (2012). Intervention for co-occurring speech and language difficulties: Addressing the relationship between the two domains. *Child Language Teaching and Therapy, 28*(1), 123–135.
Smith-Lock, K. M. (1995). Morphological usage and awareness in children with and without language impairment. *Annals of Dyslexia, 45*, 161–185.
Spada, N., & Tomita, Y. (2010). Interactions between type of instruction and type of language feature: A meta-analysis. *Language Learning, 60*(2), 263–308. https://doi.org/10.1111/j.1467-9922.2010.00562.x
Spooner, L. (2002). Addressing expressive language disorder in children who also have severe receptive language disorder: A psycholinguistic approach. *Child Language Teaching and Therapy, 18*(3), 289–313.
Stichting Siméa. (2014). *Indicatiecriteria: auditief en/of communicatief beperkte leerlingen*. Retrieved from https://www.passenderwijs.nl/wp-content/uploads/14-simea-brochure-indicatiecriteria-digitaal.pdf.
Tagarelli, K. M., Ruiz-Hernández, S., Vega, J., & Rebuschat, P. (2016). Variability in second language learning: The roles of individual differences, learning conditions, and linguistic complexity. *Studies in Second Language Acquisition, 38*(2), 293–316.
Thijs, A., Trimbos, B., Tuin, D., Bodde, M., & De Graaff, R. (2011). *Engels in het basisonderwijs [Vakrapport]. SLO: nationaal expertisecentrum leerplanontwikkeling*. http://www.slo.nl/downloads/2011/engels-in-het-basisonderwijs-vakdossier.pdf/.
Tobin, L. M., & Ebbels, S. H. (2019). Effectiveness of intervention with visual templates targeting tense and plural agreement in copula and auxiliary structures in school-aged children with complex needs: A pilot study. *Clinical Linguistics and Phonetics, 33*(1/2), 175–190.
Tomblin, B., Mainela-Arnold, E., & Zhang, X. (2007). Procedural learning in children with and without specific language impairment. *Journal of Child Language Learning and Development, 3*, 269–293.
Tribushinina, E., Dubinkina-Elgart, E., & Mak, P. (2022a). Effects of early foreign language instruction and L1 transfer on vocabulary skills of EFL learners with DLD. *Clinical Linguistics and Phonetics*, 1–18. https://doi.org/10.1080/02699206.2022.2076261
Tribushinina, E., Dubinkina-Elgart, E., & Rabkina, N. (2020). Can children with DLD acquire a second language in a foreign-language classroom? Effects of age and cross-language relationships. *Journal of Communication Disorders, 88*, Article 106049.
Tribushinina, E., Op ten Berg, Z., & Karman, S. (2022b). Facilitating positive L1 transfer through explicit spelling instruction for EFL learners with dyslexia: An intervention study. *Language Awareness, 31*(3), 351–370.
Tyler, A. (2012). *Cognitive linguistics and second language learning: Theoretical basic and experimental evidence*. Routledge.
Ullman, M. (2016). The declarative/procedural model: A neurobiologically-motivated theory of first and second language. In G. Hickok, & S. L. Small (Eds.), *Neurobiology of Language* (pp. 953–968). Elsevier.
Ullman, M., & Pierpont, M. T. (2005). Specific Language Impairment is not specific to language: The Procedural Deficit Hypothesis. *Cortex, 41*, 399–433.
Wigglesworth, G. (2002). The role of the first language in the second language classroom: Friend or foe. *English Teaching, 57*, 17–31.
Windsor, J. (2002). Contrasting general and process-specific slowing in language impairment. *Topics in Language Disorders, 22*(3), 49–61.
Zoutenbier, I., & Zwitserlood, R. (2019). Exploring the relationship between native language skills and foreign language learning in children with developmental language disorders. *Clinical Linguistics & Phonetics, 33*(7), 641–653.
Zwitserlood, R., Wijnen, F., Van Weerdenburg, M., & Verhoeven, L. (2015). 'MetaTaal': Enhancing complex syntax in children with specific language impairment-a metalinguistic and multimodal approach. *International Journal of Language & Communication Disorders, 50*(3), 273–279.