Two clinical markers for DLD in monolingual Italian speakers: what can they tell us about second language learners with DLD?

Running head: Clinical markers of DLD in L2 Italian-speaking children

Maria Teresa Guasti*[a], Michelle J. White*[b,c], Gabriele Bianco[d], Fabrizio Arosio[a], Bernard Camilleri[e], Natalie Hasson[f]

*First two authors

University of Milano-Bicocca [a]

University of Cape Town [b]

Stellenbosch University [c]

North East London NHS Foundation Trust [d]

City, University of London [e]

Independent researcher [f]

Corresponding author:

Maria Teresa Guasti

Mariateresa.guasti@unimib.it

University of Milano-Bicocca

Department of Psychology

P.zza Ateneo Nuovo 1

20126 Milano

Abstract:

A large number of children worldwide are only exposed to their L2 around 3 years of age and can exhibit linguistic behaviours that resemble those of a child with Developmental Language Disorder (DLD). This can lead to under- or over-identification of DLD in this population. This study endeavors to contribute to overcoming this problem, by determining whether two specific clinical markers used with the Italian monolingual population can also be used with early L2 acquiring children, namely clitic production and non-word repetition. Our study involved two groups of 5-year-old L2 learners of Italian from various language backgrounds; 18 children had been referred to Speech and Language Therapy (SLT) services (EL2_DLD), and 30 children were typically developing (EL2_TD). The participants completed an Italian clitic production task and a non-word repetition task based on Italian phonotactics. Data was also collected from the participants' caregivers with the ALDeQ Parental Questionnaire to obtain information about the children's L1 (Paradis, et al., 2010). Our results suggest that non-word repetition and clitic production in Italian are potentially useful for identifying L2 learners of Italian with DLD, at the age of 5 years. The repetition of non-words is highly accurate in identifying children with DLD among the participants, while clitic production is somewhat less discriminative in this sample. This study is a first step towards uncovering clinical markers that could be used to determine the presence of DLD in children acquiring their L2.

Keywords: Developmental language disorders, second language acquisition, Italian, nonword repetition, clitics

Introduction

Many children worldwide are bilingual, learning a first language (L1) at home and subsequently a second language (L2) at preschool. Encountering the L2 at preschool for the first time can be regarded as delayed exposure, which may have a long-term impact on literacy achievements (see Bonifacci &

Tobia, 2016; Kovelmann et al., 2008). In typically developing early L2 (EL2) children, the amount of language input and the age of first exposure to the L2 are generally acknowledged to be good predictors of L2 proficiency (e.g., Gathercole, 2018), yet proficiency in the L2 is also influenced by other factors (e.g. the child's L1). As a consequence of these converging factors, EL2 learners often display poor mastery of grammatical morphemes reminiscent of the error patterns that monolingual children with Developmental Language Disorder (DLD) exhibit (e.g., Håkansson, 2001; Paradis et al. 2011).

These similarities make it difficult to differentiate typically developing (TD) children from children with DLD, especially among EL2 children. Therefore, there is a risk of over- or under-identification of DLD in the L2 population. To address this, one approach, which will be pursued in this study, is to use clinical markers that have been established in monolingual children with DLD to identify potentially at-risk EL2 children. The term clinical marker refers to the linguistic behaviours that identify children as having DLD, with a high level of accuracy (sensitivity), and without erroneously including TD children (specificity). Clinical markers are often specific to families of languages. In the case of Italian, the language investigated in this study, one clinical marker is the failure to produce third person direct object (3DO) clitic pronouns. Preschool children with DLD often omit these pronouns in Italian (Bortolini et al., 2006; Guasti et al. 2016), and this failure persists into the school years (Arosio et al., 2014; Guasti et al., 2016).

A clinical marker of DLD, valid across many languages, is the diminished performance on nonword repetition (NWR) tasks (Graf Estes et al., 2007; Italian: Bortolini et al., 2006; Dispaldro et al., 2013; Dutch: de Bree et al., 2016; English: Dollaghan & Campbell, 1998; French: Thordardottir & Brandeker, 2013; Spanish: Girbau & Schwartz, 2008; Swedish: Sahlén et al., 1999; Icelandic: Thordardottir, 2008). Nonwords (NWs) are sound sequences, which adhere to the phonotactic probabilities of a given language but have no meaning in that language.

In this study, we focus on L2 Italian and on two clinical markers: clitic production and NWR. We hypothesise that clinical markers used for the monolingual population may also be valid for the

EL2 population. Before presenting the current study, the existing literature on these two clinical markers will be reviewed.

NWR tasks in monolingual and EL2 children

Several studies have established that NWR is a valid measure to identify monolingual children with DLD, yielding high levels of sensitivity and specificity, ranging from 80% to 100%. This holds true for several languages, as reported in Armon-Lotem and Meir (2016). In the EL2 population, the validity of this test is less robust than in the monolingual population, with differences among studies. A meta-analysis on the indexes for identification of DLD in Spanish-English children, (Dollaghan & Horner, 2011), found only a single study on NWR with good validity, that of Girbau and Schwartz (2008). In this study, the NWR test was carried out in the children's L1 (Spanish). It was found that the repetition of Spanish NWs discriminated between Spanish-English children with and without DLD. The test displayed good sensitivity of 82%, and very good specificity of 91%. NWR tests have displayed good validity with other language pairs. Armon-Lotem and Meir (2016) investigated the validity of the NWR using both Russian and Hebrew NWR tests with 5-to-7-year-old Russian-Hebrew children. NWR tests in both of the children's languages were accurate in identifying children with DLD, with different results for each language. Sensitivity and specificity were lower with the L1, Russian, (70% and 76% respectively) than for L2 Hebrew, (81% and 79%, respectively). Boerma et al. (2015) used two NWR tests with 5-to-6-year-old L2 Dutch speaking children; one was a Quasi-Universal (Q-U) test (Chiat 2015) and another was a language specific test. The former included NWs that respect the lexical-phonological constraints of many languages, while the latter consisted of NWs respecting the constraints of Dutch. Findings indicate that the tests could identify children with DLD in the EL2 population. The Q-U test was found to have good sensitivity (87%) and specificity (83%), while these values were lower for the language specific test: sensitivity was 77% and specificity 73%.

In Italian, the NWR test is a valid measure for identifying 3-to-5-year-old monolingual children with DLD, with 100% sensitivity and specificity (Dispaldro et al., 2013). Additionally, when

TD EL2 children (with L1 being Albanian, Romanian or a variety of Arabic) were compared to monolingual children on their NWR performance in Italian (i.e., their L2), no difference was found (Vender et al., 2016). The group of L1 Arabic-speaking children had less Italian exposure than the other groups, yet they performed as well as the other groups, in contrast with their performance on sentence comprehension. Thus, in Italian, NWR seems to be less affected by the amount of L2 exposure. This contrasts with other studies showing that performance on NWR is affected by language exposure. Sharp and Gathercole (2013) found that Welsh NWR by Welsh-English speaking children was influenced by the amount of exposure to Welsh. Similar results have been reported for Spanish-English L2 children (Summers et al. 2010). However, in partial contrast with these studies, Thordardottir and Brandeker (2013) found that the amount of exposure influences English, but not French NWR in English-French bilinguals. This discrepancy was attributed to the different phonological properties of English and French NWs: French items were phonologically less complex, with a simpler syllabic structure and stress pattern. A similar explanation was advocated by Vender et al. (2016) to explain the lack of influence of amount of exposure on NWR: Italian has a simple syllabic structure, few consonantal clusters and a regular stress pattern.

In sum, previous findings are broadly supportive of the validity of NWR for identifying children with DLD in the EL2 population. In these studies children are tested in their L1, or in both their L1 and L2, or with a Q-U NWR test. Mixed findings were observed related to the languages involved and the amount of L2 exposure.

Clitic production in EL2 children

The failure to produce 3DO clitics is a clinical marker of DLD in Italian with good sensitivity and specificity (over 80% and up to 100%) (Bortolini et al.2006; Arosio et al. 2014). It is also valid for French (e.g., Tuller et al., 2011). A study on the production of 3DO clitics in the L2 population has revealed that, after 1.5 years of exposure in immersion schools, TD 6-year-old English-French children perform as poorly as 8-year-old French monolingual children with DLD, frequently omitting

clitics (Grüter, 2005). Similarly, Chondrogianni et al. (2015) reported that Turkish-Greek children aged 7, with an average exposure to Greek of 21 months produced fewer clitics than L1 Greek-speaking children with DLD. Chondrogianni concluded that L2 Greek-speaking children have a low proficiency in the use of clitics if they have less than 3 years exposure. As for L2 Italian, Vender et al (2016) established that three groups of 5-year-old TD_EL2 children with an average exposure to Italian of 2 years, scored lower than the monolingual TD control group. However, their errors in Italian were different from those of monolingual children with DLD; while the latter group omitted clitics, the former produced an incorrect form. Vender et al., (2016) also found that among the three groups of TD_EL2 children, Arabic-speaking children produced fewer clitics than the Romanian- and Albanian-speaking children. They established that Arabic-speaking children had less exposure to Italian than the other two groups, which was reflected in their lower scores in clitic production as well as in other linguistic assessments (vocabulary and grammatical comprehension). This is consistent with the previous observation that length of exposure affects accurate clitic production.

In sum, clitic production is challenging for EL2 children, as it is for monolingual children with DLD (see also Belletti & Guasti, 2015 for a review). Overlap between the two groups may be observed in the initial stages of L2 acquisition, but some differences in terms of type of errors are observed.

The objective of the current study

We propose that EL2 children with DLD will be vulnerable, relative to EL2 children, in the same areas as monolingual children with DLD: NWR and clitic production. The present study aims to explore this by directly comparing the performance of two groups of EL2 children, one referred to clinical services and diagnosed with DLD, and one without a diagnosis of DLD. Specifically, we expect that:

(1) EL2 with DLD will repeat fewer correct NWs than EL2_TD

(2) Given a minimum of 12 months exposure to Italian, length of exposure will not affect NWR accuracy

(3) EL2 with DLD will produce fewer target clitics than EL2_TD

(4) Similarly to monolingual pre-school children with DLD, EL2 with DLD will omit clitics

(5) The production of clitics will be affected by the length of exposure to Italian

(6) The two groups of children will be further discriminated by means of a parent questionnaire about the child's L1 development

Method

Participants

Forty-eight 5-year-old EL2 children with Italian as their L2 participated in the current study. They were residing in Northern Italy in the province of Brescia. All children came from families where both parents had immigrated to Italy and were themselves L2 learners of Italian. The participants who spoke different L1s (Arabic, Romanian, Albanian, Punjabi, Serbo-Croatian, Urdu, Moldovan-Romanian, Pular, Wolof, Akan-Twi, Ghanaian and Nigerian English, Nzema, Hindi, Sinhalese, Tagalog, Russian), had been exposed exclusively or primarily to their L1 before the age of 2–3 years, and had a minimum of 12 months of exposure to Italian within a daycare/preschool setting. One group of 18 children had been referred to Speech and Language Therapy (SLT) services (EL2_DLD), and one group of 30 children (EL2_TD) were reported to be typically developing. EL2_DLD were recruited within the Health Services in Brescia and had received a diagnosis of speech and/or language impairment by a multidisciplinary team of certified speech and language therapists, audiometrists and audiologists. This diagnosis was reached through a combination of clinical judgement based on professional expertise, (which in the case of L2 learners is even more necessary given the lack of specific evaluations) and of standardised language assessments (assessing the

productive phonological repertoire, receptive/expressive vocabulary, and receptive/expressive grammar). In the absence of standardised assessments for EL2 populations, standardised tests assessing the above-mentioned language components, which were created for Italian monolinguals, were adopted using a strict cut-off of two standard deviations (2 SD) below the mean. While tests standardised on a monolingual population are not considered accurate for assessing bilingual children, here (as in clinical practice), low cut-offs on these assessments have been used to identify children whose language skills are comparable to monolingual children with severe language difficulties. Exclusionary criteria were also adopted: children with diagnosed neurological, hearing, visual, cognitive and socio-emotional deficits, as well as children diagnosed with autistic spectrum disorder, were excluded from the study.

EL2_TD children were recruited from two public preschools located in the same town within the province of Brescia. Inclusion criteria designated that there had to be no history of DLD, no other disorders or sensory problems, no reported language and cognitive difficulties and a history of typical development. Matching was based on group equivalencies (Hulley, et al., 2011). Thus, the two groups of children were matched on chronological age, length of exposure to Italian as an L2 (LE), age of first exposure to L2 (AFE), years of primary caregiver education, and non-verbal IQ. Following the results in Vender et al. (2016), LE was calculated considering the number of months the child had attended daycare/preschool up to the time of testing, removing the time spent abroad and/or away from school. Italian language input from television was not considered, due to the lack of social interaction involved in this practice (see Konishi, et al., 2014). The primary caregiver's education was chosen as a measure of socioeconomic status (SES), since it represents one of its strongest predictors (Hoff, et al., 2012). All participants of the two groups had to show typical non-verbal cognitive abilities, as measured by Raven's Coloured Progressive Matrices (CPM; Belacchi et al. 2008). None of the participants scored below -1.5 SD from the mean and no statistically significant between-group differences were found (see Table 1). The two groups could not be matched on sex; an independent t-test revealed a statistically significant difference in sex, $t(46)=3.16$, $p<.01$, with the

EL2_DLD group showing a higher male to female ratio compared to that of the EL2_TD group (5:1 vs. 0.7:1). The EL2_DLD group consisted of 15 boys and 3 girls, while the EL2_TD group consisted of 12 boys and 18 girls (see Table 1).

To ensure that sex differences did not affect the results of the present study, a preliminary statistical data analysis was conducted, in line with Peña et al. (2006). Sex (males vs. females) was entered as the between-subject independent variable, while the demographic measures were treated separately as dependent variables. Sex was not significant on any of the dependent variables. For demographic information about the two groups, as well as descriptive statistics, refer to Table 1.

PLEASE INSERT TABLE 1 HERE

The study was conducted according to the standards of the Helsinki Declaration. The relevant hospital authorities granted authorization for testing participants recruited within the Health Services, while permission for testing participants recruited within the preschools was given by the school principal. Informed consent from children's parents was obtained before the testing commenced. Children were tested in a single session, either at their schools or at SLT clinics.

Due to the heterogeneity of the migrant community in terms of geographical and language backgrounds in Italy, matching the two groups on their L1 was not possible. The participants' L1s included 15 different languages. This variety was observed in both groups. All of the children were dual language learners, except for two who were trilingual.

Materials

ALDeQ Parent Questionnaire

In order to gain insight into the participants' L1, the Alberta Language and Development Questionnaire (ALDeQ) (Paradis et al., 2010) was administered. This parent questionnaire was developed to screen the L1 development of bi/multilingual preschool-aged children. It was designed

to be non-L1 specific, and it consists of 18 questions across four sections: early milestones, current L1 abilities, activity preferences/behaviour patterns, and family history. Parents' answers are scored using rating scales, with lower scores indicating atypical development and higher scores being more consistent with typical development. For the present study, the ALDeQ was translated into Italian and administered to the parent who was the most proficient speaker of Italian. Where necessary, a family member or a family friend helped to translate the questions from Italian to the family's L1.

NWR task

A NWR task based on Italian phonotactics (Cornoldi et al., 2009) was used to test L2 Italian-speaking children. In their study, Vender et al. (2016) found that TD L2 children with different L1s (Arabic, Albanian and Romanian) did not differ from monolinguals, as far as the NWR test was concerned. In the current study, the same NWR test was adopted. This NWR task (Cornoldi et al., 2009) includes 25 items of increasing length (ranging from one to five syllable) and segmental complexity. It comprises 60 syllables in total: Twenty-four CV syllables, which is the most frequent syllable type in Italian; 19 CCV syllables, 8 CVC syllables and the remaining syllabic configurations are CCCV (4), CCVC (2), CVV (1), VC (2). All the stimuli were presented orally to the participants with the instruction to repeat the items. The responses were transcribed on-line using broad phonetic transcription, following the International Phonetic Alphabet (IPA) system. In line with prior studies (Edwards & Lahey, 1998; Gathercole et al., 1994), if a child made consistent substitution/distortion errors, these were not scored as incorrect. For the EL2_DLD group, consistent error patterns were detected using previous SLT reports, while for the EL2_TD group consistency of errors was checked with the help of preschool teachers and/or parents. Consistent with the original scoring procedure (as reported in Cornoldi et al., 2009), the participants' score corresponded to the total number of correctly repeated syllables, for a maximum of 60 syllables.

Clitic production task

The production of 3DO clitics was tested with an elicitation task adapted from Arosio et al. (2014) (reflexive clitics were not elicited). Twelve sentences were elicited which contained a 3DO-clitic. Two conditions were included: in one condition, the feminine clitic (*la*) was elicited and in the second condition the masculine clitic (*lo*) was elicited. Children were presented with two pictures on a laptop screen. While looking at the first picture, children heard the description of the event in the picture. This description was intended to provide the participants with the relevant vocabulary and topic of discourse. When the second picture appeared on the screen, participants were asked a question, which was aimed at eliciting a sentence containing a clitic pronoun. Descriptions and questions were digitally recorded by a female native speaker of Italian and played through loudspeakers connected to the laptop. An example of the elicitation material and the expected response is provided in (1). Note that the expected answer can include or omit the subject of the sentence, as Italian is a null subject language and in the context of (1), the omission of the subject is pragmatically licit.

(1) a. In questa storia c'è una signora che vuole pelare una patata
In this story, there is a lady that wants to peel a potato
b. Guarda, cosa sta facendo alla patata?
Look, what is she doing to the potato?
c. Expected answer: (la signora) la sta pelando/la pela
(the lady) it-FEM-SG is peeling/it-FEM-SG peels

The 12 experimental trials were preceded by five familiarization practice items eliciting the production of clitics; if necessary, feedback was given during the familiarization session, by providing the sentence with the clitic and asking the child to repeat it.

Response Coding

Children's responses were classified into five categories. Responses were coded as *Target* when they matched the target responses. Sentences including a wrong clitic were classified as *Wrong Form* responses. Given the target in (1c), an example of the wrong form is in (2a), in which the clitic is masculine rather than feminine. Sentences were coded as *Omission* responses when the clitic was missing, and no nominal argument was produced (see 2b). Responses were classified as *Full DP* (i.e., Determiner Phrase) when a sentence with a full nominal object rather than a clitic was produced (see 2c). This type of sentence is grammatical but pragmatically inappropriate in the context. All other responses were classified as *Other* (largely sentences that were irrelevant).

(2) a. *Lo* pela
it-MASC-SG peels

b. pela
peels

c. pela la patata
peels the potato

Results

The ALDeQ and the NWR task

The descriptive statistics for the ALDeQ is reported in Table 2 and the results of the NWR test in Figure 1. As predicted, a cursory look at the data indicates a noticeable difference between the two groups.

PLEASE INSERT TABLE 2 HERE

PLEASE INSERT FIGURE 1 HERE

An ANOVA was run with Group (EL2_DLD and EL2_TD) as a predictor and scores on the ALDeQ as the dependent variable, which yielded a significant effect of Group ($F(1,46)=57.77$, $\eta 2=0.56$, $p<0.001$), confirming our hypothesis in (6) that the EL2_DLD and the EL2_TD groups could be discriminated by means of a parent questionnaire which investigated, among others, the children's L1 development. An ANOVA with Group as a categorical predictor, Age, Age of First Exposure and Length of L2 Exposure as continuous predictors, and score on NWs as dependent variable, revealed a significant effect of Group only ($F(1,43)=176.49$, $\eta 2=0.80$, $p<0.001$). The EL2_TD children achieved scores which were almost at the ceiling score of 60. In contrast, EL2_DLD children scored well below ceiling. These findings confirm our hypotheses (1) and (2): EL2_DLD children were impaired in the repetition of NW, and length of exposure did not influence the performance of either group of children, given that they had all been exposed to Italian for a minimum of 12 months.

In order to establish the sensitivity and specificity of the NWR in classifying EL2 children with and without DLD, a Receiver Operating Characteristic (ROC) curve analysis (Zweig & Campbell, 1993) was conducted, and likelihood ratios of the optimal cut-off value were calculated. In the analysis, the number of syllables accurately repeated were used as the cut-off scores. With the ROC curve analysis, the area under the curve (AUC) measures the accuracy of a test. The AUC value for the NWR score is high, i.e., 1.00 ($p=<0.001$), indicative of a perfect test. This finding means that the NWR is highly accurate in identifying Italian-speaking children with DLD. With a cut-off criterion of 42, sensitivity and specificity are 100%. This cut-off value score has an associated negative likelihood ratio (–LR) value equal to 0 and no positive likelihood (+LR) value, since the probability of a TD child achieving a test score of ≤ 0.42 is undefined. These values classify our cut-off score as clinically informative for identifying children with DLD, since the –LR value is well below 0.10, a reference point to be considered when evaluating clinical informativeness of a test, and there is no +LR value, since the probability of a TD child achieving a test score of ≤ 0.42 is undefined (Dollaghan, 2007; Sackett et al., 1991; Sackett & Haynes, 2002).

Clitic production

Among the 18 children with DLD, six males did not take part in the clitic production test. Three of the aforementioned children did not understand the task, were not providing any verbal response or their answers were unintelligible during the practice phase. The other three children did not want to proceed with the testing. Therefore, the results of the elicitation test are based on 12 children with DLD and 30 TD children. Despite the removal of the six participants, the two groups continued to differ on the ALDeQ ($F(1,40)=34.49$, $\eta2=0.46$, $p<.001$), NWR repetition ($F(1,40) =198.7$, $\eta2=0.83$, $p<.001$) and sex ($t(40)=2.1$, $p=.04$). They did not differ with regard to chronological age, length of exposure, SES and non-verbal IQ. Furthermore, they differed as to the age of first exposure to Italian (EL2_DLD: M=43 (6.5), EL2_TD: M=39 (5.2); $t(40)=2.12$, $p=.03$), where TD children were exposed earlier than children with DLD.

Figure 2 shows the frequency of the various responses provided by the two groups of children in the clitic production task. It is evident that children in the EL2_DLD group produced fewer 3DO clitics than children in the EL2_TD group. Instead of producing a 3DO clitic, EL2_DLD children tended to omit it, to produce a post-verbal lexical full Determiner Phrase or produce other structures. Interestingly, the rate of production of clitics with wrong morphology was very low.

PLEASE INSERT FIGURE 2 ABOUT HERE

For each response category, a mixed model logistic regression analysis was conducted. By starting with a model that only included Subject and Item as random factors, we followed a stepwise inclusion procedure and evaluated which fixed factors significantly contributed to the goodness of fit of the model; this was accomplished by comparing a model including the predictor against one without it, using a $\chi2$ test (Jaeger, 2008). Then, based on z-values (Wald statistics), an estimation of the statistical significance of each predictor in the model was obtained. All statistical analyses were run using R (Lmer, version 3.4.3; R Core Team, 2017).

In the analysis of Target productions, the following factors significantly contributed to the model and were included: Group [$\chi2(1) = 11.43$, $p < .001$], Condition [$\chi2\ (1) = 6.49$, $p = .01$] and Length of L2 exposure [$\chi2\ (1) = 17.09$, $p < .001$]. Neither Age nor Sex contributed to the fit of the model and were therefore not included. As shown in Table 3a, the analysis of the production of Target clitics revealed a main effect of Group (higher production in the EL2_TD group than in the EL2_DLD group), Condition (higher production of correct *lo* clitic than correct *la* clitic) and of Length of L2 exposure (longer L2 exposure predicts production of Target structures). These results confirm our hypotheses (3) and (5).

Regarding the analysis of the production of clitics with the incorrect morphology (*Wrong Form*), Condition [$\chi2\ (1) = 7.58$, $p = .006$] and Length of L2 exposure [$\chi2\ (1) = 14.93$, $p < .0001$] significantly contributed to the model and were included, while Group did not contribute significantly and was thus not included. Both groups occasionally produced a clitic with the wrong morphology, as shown by the coefficients reported in Table 3b. In addition, children with less exposure to Italian made more errors.

In the analysis of the production of sentences with a post-verbal full Determiner Phrase (*Full DP*), Group [$\chi2\ (1) = 11.97$, $p = .0005$], Age of first exposure [$\chi2\ (1) = 4.07$, $p < .0435$] and Length of L2 exposure [$\chi2\ (1) = 5.78$, $p < .0161$] significantly contributed to the model and were included. As shown by analysis coefficients reported in Table 3c, EL2_DLD children produced more sentences with a post-verbal full DP than EL2_TD children. The production of full DP structures thus seems to decrease with longer L2 exposure. Although the Age of first exposure contributed to the model and was therefore included, its statistical effect was not significant.

The analysis of the production of sentences with omissions of the clitic (*Omissions*) showed that Group [$\chi2\ (1) = 8.2631$, $p = .0004$] and Length of L2 exposure [$\chi2\ (1) = 16.755$, $p < .0001$] significantly contributed to the model and were included. As shown by analysis coefficients reported in Table 3d, EL2_DLD children produced more sentences with omissions than EL2_TD children, as predicted by our hypothesis (4), and omission decreased with longer L2 exposure.

In the production of *Other Structures*, Group [$\chi 2(1) = 12.804$, $p = .0003$] and Length of L2 exposure [$\chi 2(1) = 6.9982$, $p = .0081$] significantly contributed to the model and were included. As shown by coefficients reported in Table 3e, EL2_DLD children produced more *Other Structures* than EL2_TD children, and *Other Structures* decreased with longer L2 exposure in both groups.

PLEASE INSERT TABLES 3A,B,C,D,E ABOUT HERE

A Receiver Operating Characteristic (ROC) curve analysis was performed and likelihood ratios of the optimal cut-off value were calculated. The number of produced Target sentences were evaluated as cut-off scores. A ROC curve for the production of the 3DO clitics is shown in Figure 3.

PLEASE INSERT FIGURE 3 ABOUT HERE

The AUC value for the target clitic production is high, i.e., 80.6 ($p=<0.001$). As an area of 0.80-0.90 indicates good accuracy, we can conclude that the production of the target clitic has a good accuracy. With a cut-off criterion of 4, sensitivity and specificity are 75% and 83%. Specificity is therefore good, while sensitivity is fair. This cut-off value has an associated +LR equal to 4.50 and -LR value equal to 0.30. These values classify our cut-off score as partially informative for identifying EL2 children with DLD, since, although the +LR value is above 3.00, the –LR value is not below 0.10, two reference points to be considered when evaluating the clinical informativeness of a test (Sackett & Haynes, 2002).

In sum, EL2_DLD children were greatly affected in the production of target clitics and their performance was clearly different from that of EL2_TD children. Rather than producing a target clitic, they omit it, produce a post-verbal DP, or an irrelevant sentence, and in that they differ from

EL2_TD children, as expected. Length of exposure to Italian is a significant predictor for both groups. Finally, failure to produce target clitics could be a marker for identifying DLD in EL2 Italian speakers, with a moderate sensitivity of 75% and a relatively high specificity of 83.3%.

Discussion

The current study demonstrated that the repetition of NWs can be excellent in identifying children with DLD in the population of EL2 children speaking Italian. The production of clitics can also be useful for the same goal, but it is less effective. In the next sections, we discuss both markers in detail.

Non-word repetition

L2 Italian speaking children were tested with a NWR test. As expected, 5-year-old EL2 children with DLD are severely affected in their repetition of Italian NWs. In this respect, they resemble Italian-speaking monolingual children with DLD at the age of 5 years. By contrast, EL2_TD children with a minimum of 12 months of exposure to Italian perform within the normal range of their Italian monolingual peers, replicating Vender et al.'s (2016) results. Note that Vender et al. used the same test as in the current study.

A cut-off point obtained for the L2 children of 42 correct syllables on the NWR task identified 100% of the EL2 children with DLD (sensitivity) as being language impaired, as well as 100% of EL2_TD children (specificity) as not impaired in language. These values are indicative of a perfect test. The same values were obtained by Dispaldro et al. (2013) with monolingual Italian-speaking children. Thus, for Italian, the NWR test might be considered as a valid measure to identify DLD, also among the EL2 population after 12 months of exposure. Our study included children with a large variety of languages (Arabic, Romanian, Albanian, Punjabi, Serbo-Croatian, Urdu, Moldovan-Romanian, Pular, Wolof, Akan-Twi, Ghanaian and Nigerian English, Nzema, Hindi, Sinhalese, Tagalog, Russian). The results for EL2_TD children displayed little variability across these different languages, suggesting that our test does not seem to be affected by properties of the L1. Our findings

are consistent with those reported by Dos Santos & Ferré (2018) on French L2 children, who showed that NWR differentiated children with and without DLD. It is also consistent with the results of Boerma et al. (2015). These authors showed that their Quasi-Universal NWR test (including CV syllables only) was clinically valid in identifying language impairment among Dutch L2-speaking children, while the language specific NWR task was not as effective.

In the current study, the EL2_TD group obtained scores within the normal range on the NWR. Other studies, however, report contradictory results. Some researchers have reported either poor sensitivity (Gutiérrez-Clellen & Simon-Cereijido, 2010, Kohnert et al., 2006) or low specificity (Windsor et al., 2010), while other authors report fair to high levels of diagnostic accuracy (Girbau & Schwartz, 2008; Thordardottir & Brandeker, 2013). The differences found between the studies can be clarified by looking at, firstly, the language tested (L1 in Girbau & Schwartz, 2008 vs. both L1 and L2 in Gutiérrez-Clellen & Simon-Cereijido, 2010, Kohnert et al., 2006 or L2 in Thordardottir & Brandeker, 2013), and secondly, the age of first exposure to the L2 (e.g., 17.9 months on average in Thordardottir & Brandeker, 2013, but unspecified in the other studies). Other important variables such as the participant age, the inclusion criteria for the language-impaired group, the amount of exposure to the L2, and the nature of the NWs contribute to the differences found between studies. For example, Messer et al. (2010) found that 4-year-old Turkish-Dutch children achieved lower scores in the recall of Dutch NWs than monolingual Dutch-speaking children. Messer et al. (2010) attributed this difference between the L1 and L2 learners to the lower knowledge of the phonotactic structure of Dutch possessed by the L2 children. Comparing the current study with the aforementioned study, one should consider that the phonotactic structure of Italian and Dutch are quite different. For example, Dutch has 19 different syllable types and many consonant clusters, while Italian has 8 syllable types and consequently fewer consonantal clusters. Thus, it is possible that the syllabic structure of Italian is less complex than that of Dutch, and this might have been beneficial for our EL2 learners (see Chiat, 2015 for the factors that affect NWR). Unlike Boerma et al. (2015), our NWs were language specific, yet the test was clinically valid. Although the majority of the

syllables in the test were CV syllables, there were also syllables with consonant clusters. However, in line with Vender et al.'s (2016) study, these consonant clusters did not negatively affect TD children's performance. Further aspects to take into consideration is that stress in Italian is regular and that no vowel reduction is observed, which may all contribute to making the phonological properties of Italian simpler than that of Dutch or English.

Finally, the length of exposure to Italian did not contribute to children's performance, after 12 months of exposure, likely because of the above-mentioned phonological properties of Italian.

Clitic production

The EL2_DLD children performed differently to the EL2_TD children, both quantitatively (fewer target responses), and qualitatively (different types of non-target responses). They did not use target 3DO clitics consistently. Moreover, scores obtained on tasks assessing the production of clitics identified 75% of the children with DLD as being language impaired and 83.3% with TD as not being language impaired. While specificity is relatively good, sensitivity is lower. Therefore, at the age tested, this test does not have a very good discriminative value. However, by looking at the responses provided, we can obtain further useful information. Firstly, the EL2_DLD group omitted more clitics than the EL2_TD group, as was anticipated based on previous literature. In this respect, they resembled 5-year-old monolingual children with DLD, as established in Bortolini et al. (2006) and Guasti et al. (2016). They also produced more sentences with a post-verbal DP and more irrelevant sentences than the EL2_TD group. These productions did not distinguish monolingual children with DLD from monolingual TD children in Guasti et al. (2016). However, the aforementioned study was the only one focusing on Italian children with DLD that had a similar coding as the current study. Irrelevant sentences and sentences with DP complements are usually removed from the analysis in other studies. Thus, further research is needed which also takes these responses into account.

The use of a DP complement rather than a clitic has been observed in 7-year old monolingual Italian-speaking children with DLD by Arosio et al. (2014). Therefore, when a more qualitative

approach is adopted, the 5-year-old EL2_DLD children in the current study appear to be both similar (omission) and different (DP and irrelevant sentences) from monolingual children with DLD.

Length of exposure was beneficial to both groups of children in the sense that target productions increased with longer time spent in an Italian language environment. This finding is in line with Vender et al.'s (2016) study on EL2_TD children.

EL2_TD children in Vender et al.'s study had L1s in which clitics are used (Albanian, Arabic, Romanian). These L1s were also present in our sample, however there were other L1s without clitics. Unfortunately, given the high number of different languages spoken by the children in our study, it is difficult to establish whether the use of clitics in L2 is influenced by the presence of clitics in L1.

Lastly, a comparison of the NWR task and the clitic task should be considered. While an effect was found of the amount of exposure to Italian in the clitics task, this effect was not observed in the NWR task. It is possible that to achieve competence in the production of clitics, a greater amount of Italian exposure is needed than that which is needed to accurately repeat NWs, as argued by Chondrogianni (2008).

General conclusions

Our aim was to investigate whether clinical markers established for the monolingual population can offer some insight into detecting DLD in EL2 children. The current results indicate that these clinical markers have the potential to be useful in children acquiring Italian as their L2. At 5 years of age, and after a minimum of 12 months of exposure to the L2, the NWR task is extremely sensitive in identifying children with DLD among the EL2 population, while clitic production is less discriminative. However, EL2 Italian children with and without DLD differ in their responses, both quantitatively and qualitatively.

Future research is needed to establish the validity of these insights. A longitudinal study design which follows bilingual children (both referred and non-referred) would offer clearer insights into the predictive value of the specificity and sensitivity of these clinical markers. In addition, once

difficulties learning Italian have been identified, bilingual children should be further assessed in Italian as well as in their first language, as recommended by the International Association of Logopedics and Phoniatrics (IALP, Fredman, 2006).

Statement of interest

The authors report no conflicts of interest.

References

Armon-Lotem, S., & Meir, N. (2016). Diagnostic accuracy of repetition tasks for the identification of specific language impairment (SLI) in bilingual children: evidence from Russian and Hebrew. *International journal of language & communication disorders*, 51: 715-731.

Arosio, F., C., Branchini, L., Barbieri & M. T., Guasti (2014) Persistency of direct object clitic omission in Italian school age children with SLI. *Clinical linguistics and Phonetics.* 1-25.

Belletti, A., & Hamann, C. (2004) On the L2/bilingual acquisition of French by two young children with different source languages. In P. Prévost and J. Paradis (eds) The Acquisition of French in Different Contexts: Focus on Functional Categories (pp. 147-176). Amsterdam: John Benjamins.

Bishop, D. V., North, T. & Donlan, C. (1995), Genetic Basis of Specific Language Impairment: Evidence from a Twin Study. *Developmental Medicine & Child Neurology*, 37: 56-71.

Bonifacci, P. & Tobia, V. (2016) Crossing barriers: Profiles of reading and comprehension skills in early and late bilinguals, poor comprehenders, reading impaired, and typically developing children. *Learning and Individual Differences*, 47, 17-26.

Bortolini, U., Arfé, B., Caselli, C. M., Degasperi, L., Deevy, P., & Leonard, L. B. (2006). Clinical markers for specific language impairment in Italian: The contribution of clitics and NWR. *International Journal of Language & Communication disorders,* 41, 695-712.

Boerma, T., Chiat, S., Leseman, P., Timmermeister, M., Wijnen, F., & Blom, E. (2015). A quasi-universal nonword repetition task as a diagnostic tool for bilingual children learning Dutch as a second language. *Journal of Speech, Language, and Hearing Research*, *58*(6), 1747-1760.

Chiat, S. (2015). Nonword repetition. In Armon-Lotem, S, de Jong, J. & N. Meir (eds), *Assessing multilingual children: Disentangling bilingualism from language impairment*, Bristol: Multilingual Matters, 123-148.

Chondrogianni, V. (2008) The acquisition of determiners and clitic pronouns by child and adult L2 learners of Greek. Unpublished doctoral dissertation, University of Cambridge.

Chondrogianni, V., Marinis, T., Edwards, S., & Blom, E. (2015). Production and on-line comprehension of definite articles and clitic pronouns by Greek sequential bilingual children and monolingual children with specific language impairment. *Applied Psycholinguistics,* 36, 1155-1191. doi:10.1017/S0142716414000101

Cornoldi, C., Miato, L., Molin, A., & Poli, S. (2009). *PRCR-2. Prove di prerequisito per la diagnosi delle difficoltá di lettura e scrittura.* Giunti Organizzazioni Speciali.

de Bree, E., Rispens, J., & Gerrits, E. (2007). NWR in Dutch children with (a risk of) dyslexia and SLI. *Clin. Linguist. Phonet, 21*, 935–944.

Dispaldro, M., Leonard, L.B. and Deevy, P. (2013) Real-word and nonword repetition in Italian-speaking children with specific language impairment: A study of diagnostic accuracy. *Journal of Speech, Language, and Hearing Research* 56 (1), 323-336.

Dollaghan, C.A., & Horner, E.A. (2011). Bilingual Language Assessment: A Meta-Analysis of Diagnostic Accuracy. *Journal of Speech, Language, and Hearing Research,* 54, 1077-1088. doi:10.1044/1092-4388(2010/10-0093)

Dos Santos, C., & Ferré, S. (2018). A nonword repetition task to assess bilingual children's phonology. *Language Acquisition, 25*(1), 58-71. Dollaghan, 2007

Edwards, J., & Lahey, M. (1998). Nonword repetitions of children with specific language impairment: Exploration of some explanations for their inaccuracies. *Applied Psycholinguistics, 19*, 278–309.

Fredman, S. (2006), Transformation or Dilution: Fundamental Rights in the EU Social Space. European Law Journal, 12: 41-60.

Gathercole, S. E. (2006). Nonword repetition and word learning: The nature of the relationship. *Applied Psycholinguistics,* 27, 513-543.

Gathercole, S. E., Willis, C., Baddeley, H., & Emslie, A. D. (1994). The Children's Test of Nonword Repetition: A test of phonological working memory. *Memory, 2,* 103–127.

Girbau, D., & Schwartz, R. G. (2008). Phonological working memory in Spanish-English bilingual

children with and without specific language impairment. *Journal of Communication Disorders*, 41, 124–145.

Graf Estes, K., Evans, J., & Else-Quest, N. (2007). Differences in nonword repetition performance of children with and without Specific Language Impairment: A meta-analysis. *Journal of Speech, Language, and Hearing Research*, 50, 177-195.

Grüter, T. (2005) Comprehension and production of French object clitics by child second language learners and children with specific language impairment. *Applied Psycholinguistics*, 26, 363-391.

Grüter, T., & Crago, M. (2012). Object clitics and their omission in child L2 French: The contributions of processing limitations and L1 transfer. *Bilingualism: Language and Cognition*, 15, 531–549.

Guasti, M.T., Palma, S., Genovese, E., Stagi, P., Saladini, G. & Arosio, F. (2016). The production of direct object clitics in pre-school and primary school-aged children with specific language impairments. *Clinical linguistics and Phonetics*, 30, 663-678.

Gutiérrez-Clellen, V. F., & Simon-Cereijido, G. (2010). Using nonword repetition tasks for the identification of language impairment in Spanish-English-speaking children: does the language of assessment matter? *Learning Disabilities Research & Practice*, 25, 48-58.

Håkansson, G. (2001). Tense morphology and verb-second in Swedish L1 children, L2 children and children with SLI. *Bilingualism: Language and Cognition*, 4, 85-99.

Hoff, E., Laursen, B., & Bridges, K. (2012). Measurement and model building in studying the influence of socioeconomic status on child development. In Lewis, M., Mayes, L. (Eds.), *The Cambridge handbook of environment in human development*, 590-606. Cambridge: Cambridge University Press.

Hulley, S. B., Cummings, S. R., Browner, W. S., Grady, D. G., & Newman, T. B. (2011). *Designing Clinical Research*. Lippincott Williams & Wilkins.

Kohnert K., Windsor J., & Kim, D. (2006). Do language-based processing tasks separate children

with Language Impairment from typical bilinguals? *Learning Disabilities Research and Practice*, 21, 19–29.

Konishi, H., Kanero, J., Freeman, M. R., Golinkoff, R. M., & Hirsh-Pasek, K. (2014). Six Principles of Language Development: Implications for Second Language Learners. *Developmental neuropsychology*, *39*(5), 404-420.

Kovelman, I., Baker, S. A., & Petitto, L. A. (2008). Age of first bilingual language exposure as a new window into bilingual reading development. *Bilingualism: language and cognition*, *11*(2), 203-223.

Messer, M. H., Leseman, P. P. M., Boom, J., & Mayo, A. Y. (2010). Phonotactic probability effect in nonword recall and its relationship with vocabulary in monolingual and bilingual preschoolers. *Journal of Experimental Child Psychology*, 105, 306-323.

MIUR (Italian Ministry of Education, University and Research) & ISMU (Foundation for Initiatives and Studies on Multi-Ethnicity) (2015). Alunni con cittadinanza non italiana. Tra difficoltà e successi. Rapporto nazionale A.s. 2013/14. Retrieved July, 1, 2015, from http://www.istruzione.it/allegati/2015/Rapporto_alunni_cittadinanza_non_italiana_2013_14.pdf

Paradis, J., & Crago, M. (2000). Tense and temporality: similarities and differences between language-impaired and second-language children. *Journal of Speech, Language and Hearing Research*, 43, 834-847.

Paradis, J., Emmerzael, K., & Sorenson Duncan, T. (2010). Assessment of English Language Learners: Using Parent Report on First Language Development. *Journal of Communication Disorders, 43*, 474-497.

Paradis, J., Genesee, F., & Crago, M. (2011). *Dual language development and disorders: A handbook on bilingualism and second language learning* (2nd Edition). Baltimore, MD: Brookes.

Prévost, P. Tuller, L., Scheidnes, M., Ferré, S., & Haiden, M. (in press). Computational complexity effects in the acquisition of wh-questions in child l2 French. In L. Dominguez & P. Guijarres-

Fuentes (eds.), *Selected proceedings of the Romance Turn III*, Cambridge: Cambridge Scholar Publisher.

R Core Team (2017). R: A language and environment for statistical computing. *R Found. Stat. Comput. Vienna, Austria. URL http://www. R-project. org/., page R Foundation for Statistical Computing*.

Sackett, D.L., Hayes, R.B., Guyatt, G.H., & Tugwell, P. (1991). Clinical epidemiology — a basic science for clinical medicine, second edition. New York: Little, Brown. Edinburgh: Churchill Livingstone.

Sahlén, B., Reuterskiold-Wagner, C., Nettelbladt, U. and Radeborg, K. (1999) Nonword repetition in children with language impairment - pitfalls and possibilities. *International Journal of Language and Communication Disorders 34* (3), 337-352.

Sharp, K.M. & Gathercole, V.C.M. (2013) Can a novel word repetition task be a language-neutral assessment tool? Evidence from Welsh-English bilingual children. *Child Language Teaching and Therapy*, 29, 77-89.

Summers, C., Bohman, T.M., Gillam, R.B., Pen a, E.D. and Bedore, L.M. (2010) Bilingual performance on nonword repetition in Spanish and English. *International Journal of Language and Communication Disorders*, 45, 480-493.

Topbas, S. & Kacar, D.K. (2013) A comparative study: The implementation of a Turkish Nonword Repetition Test and Quasi-Universal Nonword Repetition Test with monolingual Turkish children with SLI. Paper presented at COST Action IS0804 meeting held in Lisbon, 13-15 February 2013.

Tuller, L., Delage, H. and Monjauze, C. (2011) Clitic pronoun production as a measure of atypical language development in French. *Lingua*, 121, 423-441.

Vender, M., Garraffa, M., Sorace, A., & Guasti, M. T. (2016). How early L2 children perform on Italian clinical markers of SLI: A study of clitic production and nonword repetition. *Clinical linguistics & phonetics*, *30*(2), 150-169.

Windsor, J., Kohnert, K., Lobitz, K. F., & Pham G.T. (2010), Cross-language nonword repetition by bilingual and monolingual children. *American Journal of Speech-Language Pathology*, 19, 298–310.

Youden, W.J. (1950). Index for rating diagnostic tests. *Cancer, 3,* 32–35.

Zweig, M.H. & Campbell, G. (1993) Receiver-Operating Characteristic (ROC) Plots: A Fundamental Evaluation Tool in Clinical Medicine. *Clinical Chemistry, 39*, 561-577.

Figure Captions

Figure 1. Number of accurately repeated syllables (max=60) in the non-words repetition task by the EL2_DLD and EL2_TD groups.

Figure 2. The distribution of responses by the EL2_DLD and EL2_TD groups in the clitic production task.

Figure 3. ROC curve for the production of the third person DO clitics.