Non-word repetition in bilingual children: the role of language exposure, vocabulary scores and environmental factors

Non-word repetition in bilingual children: the role of language exposure, vocabulary scores and environmental factors

Gianmatteo Farabolini[ab], Pasquale Rinaldi[a], Maria Cristina Caselli[a], Alejandrina Cristia[b]

[a]*Institute of Cognitive Sciences and Technologies, National Research Council, Rome, Italy*

[b]*Départment d'études cognitives Laboratoire de Sciences Cognitives et Psycholinguistique, Paris, France*

Corresponding author:
Gianmatteo Farabolini
Università Politecnica delle Marche
Department of Clinical and Experimental Medicine
Via Conca, 71
Ancona, 60121, Italy
e-mail:
g.farabolini@pm.univpm.it
gianmatteo.farabolini@gmail.com

Number of words: 13435

Non-word repetition in bilingual children: the role of language exposure, vocabulary scores and environmental factors

Abstract

Assessing language development in bilingual children is challenging in geographical areas where bilinguals have different native languages. Lexical development measures are often used as a starting point to study linguistic abilities in bilingual children. Non-word repetition (NWR) has been found to be very informative in detecting variation. The present study contributes to the broader research aim of documenting bilinguals' language skills. In a sample of 19 Italian-speaking bilingual children with different native languages, correlations among performance on an Italian-like NWR and receptive vocabulary score (Italian PPVT-R), cumulative exposure, age of first exposure to Italian, current Italian exposure, maternal education, parental concerns and vocabulary in toddlerhood (MB-CDI) were calculated. NWR performances correlated with PPVT-R and parental concerns, but not with maternal education and language exposure measures. Neither NWR scores nor PPVT-R scores were related to Italian vocabulary size in toddlerhood (MB-CDI). We integrate our results with those of others and discuss the advantages and disadvantages of administering NWR to bilingual children, and more generally how to perform early bilingual language assessments.

Keywords: non-word repetition, bilingualism, language exposure, parental concerns, receptive vocabulary, bilingual language assessment.

Introduction

Monolingual children vary greatly in the speed and manner in which they acquire language (Frank et al., 2017). Developmental paths of bilinguals appear even more diverse, potentially due to variation in important factors such as the quantity and quality of language exposure (Carroll, 2016; Gatt & O'Toole, 2016). Some heterogeneity is considered to be within the range of what is typical, but other variation can be a sign of atypical development. The present study contributes to the broader research aim of documenting bilinguals' language skills. More specifically, we relate performance in a non-word repetition (NWR) task

administered at age 4 years to previous and concurrent vocabulary, language exposure measures, parental concerns, and maternal education. Additionally, we explore the predictive value of monolingual vocabulary assessed in toddlerhood on a NWR task and a monolingual vocabulary test.

Detecting atypical language paths

Different authors underlined the need for early identification of atypical language development in order to promote early language intervention and improve language and overall outcomes in children with persistent language difficulties (Chilosi et al., 2019; Fisher, 2017; Lüke et al., 2016; St Clair et al., 2019). In monolingual children, before the age of 3 years, atypical language developmental paths are usually detected through low performance in expressive vocabulary assessments, often relying on parental reports (Fenson et al., 2007; Rescorla, 1989). One such instrument is the MacArthur-Bates Communicative Development Inventory (MB-CDI), a parental questionnaire often used to identify toddlers with an expressive delay, who are usually called late talkers (Marini et al., 2017; Desmarais et al., 2008). Following Bishop et al. (2017), we will talk about language difficulty rather than delay. After the age of 4 years, children with language difficulties in toddlerhood can evolve in several ways. Some of them receive a diagnosis of developmental language disorder (DLD), showing persistent language difficulties in one or more language domains (Bishop et al., 2017). Others, called *late bloomers,* bridge their linguistic gap and show similar language performance as their peers without a history of language difficulties (Desmarais et al., 2008; Reilly et al., 2018).

It is difficult to transpose this approach to bilinguals, for several reasons. First, assessing bilinguals in one language and comparing them to monolinguals is not fair, because such instruments capture vocabulary of monolinguals better than that of bilinguals. Second,

norms often do not reflect the bilingual population, since tests are not typically normed on bilingual samples (Rinaldi et al., 2019; Rescorla, 1989).

ASHA and IALP recommend assessing both languages in bilingual children (ASHA, 2004; IALP, 2011). However, it is very difficult to assess language development in all the target languages the child is exposed to during her lifespan. First of all, clinicians have to be proficient in all the different native languages (i.e., languages spoken at home) spoken by bilinguals in a geographical area. Then they are called to evaluate children on all the languages the children are exposed to, using assessment tools with appropriate norms or statistically controlling for differences in age at onset and amount of input. Moreover, it has to be underlined that acquiring data in the languages spoken by bilingual children is mainly feasible in geographical areas with a homogenous bilingual population. Indeed, in some US regions, such as Florida and California, some bilingual assessment procedures for bilingual language development have been created. In countries, like Italy, where bilingual children are exposed to a wide variety of native languages in addition to a common major language, e.g., Italian in Italy and in the current study, some of the stumbling blocks for a bilingual language assessment are: a lack of culturally sensitive assessment tools, a dearth of bilingual speech and language therapists, and reduced financial aids (Armon-Lotem, de Jong & Meir, 2015). Finally, there is no method for measuring *absolute* quantities of exposure in one language (i.e., there is no way of asking the parent how much people talk to their child and expect them to estimate this accurately).

It is likely that bilinguals vary more than monolinguals do because key causal factors affecting language development in the latter vary to an even greater extent among bilinguals. One such factor is language exposure (e.g.: Carroll, 2016; Hoff et al., 2012). Specifically, evidence supports the claim that quantity and quality of language exposure of each language

spoken by bilinguals affect their lexical and grammatical development (Gatt & O'Toole, 2016; Hoff et al., 2012).

It turns out that, despite its potential weaknesses, assessing language development with vocabulary checklists in the major language remains the most viable tool for clinicians interested in serving bilingual and multilingual children younger than 3 years, in geographical areas, like Italy, where there are different cultural and linguistic minorities and children from migrant and immigrant parents have diverse native languages (Onofrio, Rinaldi & Pettenati, 2012).

Regarding bilingual language assessment after the age of 3, a recent research network attempts to assess bilingual language-general (as opposed to language-specific) development, with the main aim to optimally exclude the impact of prior language exposure and language-specific proficiency of bilingual children, in order to minimize the impact of bilingual status (Armon-Lotem, de Jong & Meir, 2015). This is particularly useful in geographical areas where children are exposed to one major language and to a variety of other native languages. For instance, different non language-specific (so-called crosslinguistic) assessment tools were developed with the aim of identifying atypical language development in bilingual children: a receptive and expressive lexical task (Haman, Łuniewska & Pomiechowska, 2015), a narrative retelling (Gagarina et al., 2012), and a crosslinguistic NWR list (Chiat & Polišenská, 2016). To our knowledge, some adaptations are available for different languages for the crosslinguistic lexical (Haman, Łuniewska & Pomiechowska, 2015), narrative (Gagarina, Bohnacker & Lindgren, 2019) and NWR (Boerma et al., 2015; dos Santos & Ferré, 2016; Grimm et al., 2014) tests, but no adaptations have been published for bilingual Italian-speaking populations.

NWR and individual variation among bilingual children

NWR tasks have been found to be sensitive in identifying atypical language development in childhood indexing fine individual differences in working memory, phonological knowledge, and/or lexical skills, among others (Armon-Lotem, de Jong & Meir, 2015; Marini et al., 2015). In it, children hear a nonsense word (i.e., an item that has no meaning in the target language, but it sounds like a word in that language) and have to immediately repeat it.

Working memory is a complex cognitive function, constituted by a central executive system that controls and manages the information flow that has to be treated, a visuo-spatial sketchpad and an articulatory loop. In the articulatory loop, there is a phonological short-term memory component, which is responsible for controlling articulatory processes selecting linguistic material that has to be maintained for a brief period for language treatment (Baddeley, 1986). In NWR tasks, children cannot use any well-learned word production routines and must instead piece together the item using phonological audio-articulatory representations (Baddeley, 1986; Dispaldro, Leonard & Deevy, 2013). Thus, the nature of NWR tasks allows us to explore the functioning of phonological short-term memory components, without primarily involving the linguistic abilities of a target language (Armon-Lotem & Meir, 2016; Munson, Kurtz & Windsor, 2005). Recent research designs use crosslinguistic nonwords in order to maximally reduce the influence of lexical and sub-lexical skills on NWR tasks (Armon-Lotem, de Jong & Meir, 2015). Beyond research designs, NWR is considered a neuropsychological task since it mainly involves phonological short-term memory processes, with different degrees of influence from language-specific proficiency depending on the nature and features of the NWR stimuli (Munson, Kurtz & Windsor, 2005). Focusing on monolingual Italian samples, three relevant studies showed that NWR distinguishes monolingual children with typical language development from those with atypical language development: NWR performance was significantly different in children

with versus without a developmental language disorder (Estes, Evans & Else-Quest, 2007), and with versus without a history of expressive language difficulties (Marini et al., 2017; D'Odorico et al., 2007).

As discussed briefly above, NWR lists can be developed following different research designs: there are language-like nonwords that are built with lexical and phonological cues of a target language, or stimuli elaborated following phonological properties shared across two or more different languages, in which case they are called non-wordlike (Thordardottir, 2017), quasi-universal (Armon-Lotem & Meir, 2016; dos Santos & Ferré, 2016) or crosslinguistic (Chiat & Polišenská, 2016) nonwords. Crosslinguistic stimuli are sometimes recommended over language-specific ones (Thordardottir, 2017), especially with bilinguals. The intuition is that, by building items that are not specific to a given language, strong phonological skills in any of the languages spoken by the child may be helpful for such a task (Boerma et al., 2015; Chiat & Polišenská, 2016). It should be noted, however, that it is impossible to build a truly universal list because languages' phonologies can vary in mutually incompatible ways. For example, it is impossible to build a 10- or 20-item list that is phonologically legal but lexically meaningless in languages that are very different in their phonological properties. Consider the case of a language that requires closed or heavy syllables (such as English), together with a language that disallows closed syllables (such as Mandarin). This is a particularly clear example, but similar problems will emerge when trying to construct stimuli for two tonal languages that have different rules governing the co-occurrence of tones and syllable structures or tones and phonemes, languages that have incompatible phonotactics, and even languages whose phonological inventory overlaps little.

Another aspect of NWR stimuli that can affect performance is length. Some authors have found significant correlations between the length of NWR stimuli (in, e.g., number of syllables) and children's NWR performance (Marini et al., 2017; Windsor et al., 2010):

longer NWR stimuli are more difficult to repeat than shorter stimuli both for monolingual (Coady & Evans, 2008; Munson, Kurtz & Windsor, 2005; Petruccelli, Bavin & Bretherton, 2012) and bilingual (Gibson et al., 2015; Summers et al., 2010) children. This effect suggests that longer non-words impose a greater load on phonological working memory, which (as summarized above) might be enhanced by lexical knowledge (Coady & Evans, 2008).

Evidence on NWR tasks to identify bilingual children with atypical language development is controversial. Some authors found that certain NWR lists were an accurate clinical marker to detect DLD as well as to identify early language difficulty (Guiberson & Rodríguez, 2015) using language-specific nonwords in one language (Girbau & Schwartz, 2008; Saiegh-Haddad & Ghawi-Dakwar, 2017) or two (Gibson et al., 2015); other work uses crosslinguistic nonwords (Armon-Lotem & Meir, 2016; dos Santos & Ferrè, 2016). Others have found that English-like NWR was not useful to identify Spanish-English bilinguals with DLD (Gutiérrez-Clellen & Simon-Cereijido, 2010; Kohnert, Windsor & Yim, 2006). Nevertheless, in situations where neither crosslinguistic nor language-specific NWR lists for all the languages spoken by children are available, bilinguals can only be assessed using a language-specific NWR task created to assess NWR performance in the major language (in our study, Italian), keeping in mind that word acquisition processes and lexical abilities in that language can influence NWR performance (Armon-Lotem & Meir, 2016; dos Santos & Ferré, 2016).

Factors related to NWR performance

As noted briefly above, NWR performance likely reflects a host of factors varying across individuals, factors which may to some extent be correlated among them. It is beyond the scope of this paper to perform a systematic review on all possible predictors of NWR performance, but we would like to close this overview with a synthesis on some key factors

relevant to the present study: vocabulary score, language exposure, socioeconomic status, and parental concerns.

Different theories suggest lexical abilities might be related to NWR performance in both monolingual (Munson, Kurtz & Windsor, 2005; Stokes, Moran & George, 2013) and bilingual (see Appendix A) children. One of the main arguments is that in childhood children have already started developing their receptive and expressive vocabulary: lexical and sub-lexical components involved in word acquisition mechanisms could enhance cognitive and linguistic processes enrolled during NWR tasks (Coady & Evans, 2008). Mixed results are reported on the correlation between NWR and vocabulary measures, with the majority of studies reporting moderate to strong results. While significant correlations were found in samples of monolingual (D'Odorico et al., 2007; Marini et al., 2017; Stokes, Moran & George, 2013) and bilingual (Guiberson & Rodríguez, 2015) children with and without early language difficulties, some authors found non-significant correlations among monolingual toddlers (Stokes, Moran & George, 2013) and children (McBride-Chang et al., 2005), as well as in bilingual children (Park & Schwarz, 2016). Finally, mixed evidence has been found for language-like (e.g., Marini et al., 2017; Stokes et al., 2013) and non-wordlike (e.g.: Chiat & Polišenská, 2016; Huls, 2017) stimuli, with a longitudinal study reporting consistent correlations between receptive vocabulary and language-like NWR across ages, but mixed evidence for non-wordlike stimuli (White, 2020). In previous literature, correlations on bilingual samples are often calculated assessing NWR accuracy and lexical scores within the same target language, e.g., English NWR and English lexical scores (Core et al., 2017; Duncan & Paradis, 2016). In a publication with a sample size similar to our study, authors found small to moderate effects and non-significant results on Korean-English bilingual children (Park & Schwarz, 2016).

As mentioned previously, exposure to a given language varies across bilingual children. Although there is disagreement as to how to measure language exposure optimally (Hoff, 2020; Thordardottir, 2017), it is likely that this factor is crucial for language acquisition (Carroll, 2016; Hoff et al., 2012; Place & Hoff, 2011). Three language exposure measures are frequently used to study its association with NWR, namely age of first exposure, current exposure, and cumulative exposure. Regardless of the definition, results on whether NWR performance correlates with language exposure among bilinguals are mixed (see Appendix B), as we illustrate next.

Both significant (Thordardottir & Juliusdottir, 2013; Li'el, 2017) and non-significant (de Almeida et al., 2017) correlations have been found between NWR performance and age of first exposure to a target language, while others found both significant and nonsignificant associations in the same sample (Huls, 2017; Summers et al., 2010). For example, in a study on 47 Spanish-English bilinguals living in the US, results revealed that the age of first exposure to English was not significantly related to English-like and crosslinguistic NWR scores, but significantly related to a composite score of English- and Spanish-NWR performance (Huls, 2017).

Current exposure is the amount of exposure to a target language calculated in~~on~~ a short period just before the assessment. Current exposure was significantly correlated with NWR accuracy in a sample of 45 Welsh-English preschoolers (Sharp & Gathercole, 2013), while in a sample of 135 Spanish-English toddlers the association between current Spanish exposure and English-NWR performance approached significance, but current Spanish exposure was not significantly related to Spanish-NWR accuracy (Core et al., 2017).

Finally, cumulative exposure is measured in diverse ways across papers. For example, one definition is based on the amount of exposure to a target language calculated in daily waking hours (e.g., Parra, Hoff & Core, 2011); another relies on an index considering

settings, speakers, and speakers' speech features (Thordardottir & Brandeker, 2013); a third considers the length of institutional exposure to the major language (Duncan & Paradis, 2016). The association between cumulative exposure and NWR performance was significant in some studies (Duncan & Paradis, 2016; Gibson et al., 2015; Thordardottir, 2017) and not significant in others (Altman et al., 2014; Abed Ibrahim & Fakete, 2018; Parra, Hoff & Core, 2011); yet others report both significant and non-significant correlations in the same sample (Brandeker & Thordardottir, 2015; Thordardottir & Brandeker, 2013).

These three implementations (age of onset, current exposure, cumulative exposure) presumably measure different aspects of the experience and its effect, and therefore we report below on all three. Similarly to NWR and receptive vocabulary, in previous literature it is common to calculate correlations between NWR performance of nonwords developed using language-specific features and exposure to that same language (e.g., Spanish-NWR performance and Spanish language exposure in Core et al., 2017; but see also de Almeida et al., 2017; Thordardottir, 2017). Moreover, some studies analyzed the relationship between language exposure and NWR performance comparing monolinguals against bilinguals: they found both significant (e.g., Engel de Abreu et al., 2013; Haman et al., 2017) and non-significant (Chiat & Polišenská, 2016; Paradis, Emmerzael & Duncan, 2010) correlations.

Other variables might be related to NWR performance besides vocabulary development and language exposure. First, socioeconomic status (SES) seems to be a relevant variable among environmental factors that influence the cognitive and linguistic development in infants (Meir & Armon-Lotem, 2017). SES is indexed by parental occupation (Marini et al., 2017), household and neighborhood information (Li'el, 2017) or, more commonly, parental education (Balladares, Marshall & Griffiths, 2016; Meir & Armon-Lotem, 2017; Seeff-Gabriel, Chiat & Roy, 2008). Mixed results are found in the

literature on the association between NWR performance and SES; more specifically, mixed evidence has been found considering NWR research design. While non-significant associations were reported between SES and non-wordlike stimuli (Antonijevic et al., 2019; Boerma et al., 2015; Chiat & Polišenská, 2016), both significant (Meir & Armon-Lotem, 2017) and non-significant (Balladares, Marshall & Griffiths, 2016; Huls, 2017) correlations have been found using language-like nonwords.

Second, while parental concerns have been found informative on both monolingual and bilingual language development (Bedore et al., 2011; Boerma et al., 2015, Tuller et al., 2018), few studies explore the association between NWR score and parental concerns. A recent study evaluated bilingual children using a multiple risk factors index and found parental referrals were related to language development (typical vs atypical language development) above and beyond other risk factors in French schoolers with different non-French native languages (de Almeida et al., 2017).

The current study

The current study aims at providing one new data point on language assessments of young bilinguals. In a sample of Italian-speaking bilingual children, we collected NWR performance, estimated Italian receptive vocabulary, language exposure to Italian, parental concerns about language development and maternal education. Additionally, given the paucity of evidence on bilingual lexical development from toddlerhood to childhood (Hoff et al. 2014a), we explored whether Italian expressive vocabulary in toddlerhood predicted later differences in performing an Italian-like NWR task in childhood among bilinguals (both in terms of correlations and via a subgroup analysis), as well as receptive vocabulary score on a picture pointing test.

We aimed to contribute data to these questions:

(1) What is the association strength between NWR performance and receptive vocabulary, language exposure measures, parental concerns and maternal education?
(2) What is the association strength across monolingual assessments in the major language (Italian), specifically, between vocabulary size in toddlerhood (i.e., at about 21.7 – 31.4 months of age), NWR accuracy and vocabulary score in childhood?

In general terms, the following results and interpretations could have been predicted, although evidence for most of these predictions is mixed, as summarized above. Regarding question (1), a significant correlation between receptive vocabulary and NWR would support the hypothesis that language-like nonwords are influenced by lexical and sub-lexical cues. Language exposure measures should be associated with NWR performance, since language-like nonwords performance might be affected by language exposure. Similarly, language-like nonwords, which should be potentially affected by language-specific development, could be associated with maternal education, which has been reported to be associated with language development. Finally, parental concerns could be correlated to NWR performance as parental concerns have been found to be a reliable preliminary index of children's language development. As for (2), we expected correlations among these metrics, which are all indices of children's Italian skills.

Materials and Methods

Participants

Nineteen children (9 girls) participated in this study. They were between 3.8 and 5.6 years of age (mean age 4.4 years) when NWR and background information were collected. All children were born in Italy from families in which neither parent spoke Italian natively or one of the parents had a native language different from Italian. As for children whose both parents spoke a language other than Italian natively, two had parents who spoke Spanish, two

Romanian, and the remaining children had Arabic-, Creole-, Polish-, Tagalog- and Ukranian-speaking parents. As for the participants having one parent who spoke a language other than Italian, three children were exposed to French, two to English, two to German, and one to each of the following: Croatian, Moldavian and Romanian. Children can thus be considered as bilingual since they were exposed to Italian and one more language early on in their life (age of Italian onset ranged from birth to 22 months, mean 6.42 months, depicting the bilingualism heterogeneity). This heterogeneity does not decrease methodological rigor and it reflects bilingualism complexity (Gatt, O'Toole & Haman, 2015). Cumulative exposure, current exposure and age of Italian onset were computed as language exposure measures to quantify bilingual language exposure. All of the children were born full-term (at least 37 weeks gestational age) and had an uneventful birth. No parent reported suspicions of cognitive, neurosensory, or psychiatric disorders. Parents of children had filled in the Italian Words and Sentences Short Form of the MacArthur-Bates CDI (MB-CDI) when children were younger than 3 years of age (Onofrio, 2017). Furthermore, in order to explore the predictive value of MB-CDI scores in toddlerhood on NWR performance across word length in childhood, we divided children in two sub-groups on the basis of having obtained a MB-CDI score either below or above the 10th percentile (Desmarais et al., 2008; Fenson et al., 2007); eight children (4 girls) obtained a MB-CDI score lower than the 10th centile in toddlerhood (see Appendix F).

Instruments and procedure

Children were evaluated at either the Language and Communication across Modalities laboratory of the Institute of Cognitive Sciences and Technologies in Rome or their home. Experimental data were collected between February 2018 and May 2018. A speech and language therapist (the first author) administered the tests in a quiet room, and each child was

assessed individually, which took 45 minutes. All testing sessions were videotaped, transcribed, and analysed.

Children were assessed with two tests based on direct observation: (a) the NWR test (Dispaldro, Leonard & Deevy, 2013) and (b) the Italian version of the Peabody Picture Vocabulary Test (Stella, Pizzoli & Tressoldi, 2000). Parents of children were administered an Italian adaptation of a semi-structured interview to measure language exposure, maternal education and parental concerns (see Appendix C).

The NWR test (Dispaldro, Leonard & Deevy, 2013) was used to assess children's repetition of nonwords. In that work, 24 two- to four-syllable nonwords (8 stimuli for each word length) had been designed following Italian phonological constraints. Those authors specified that they built stimuli from real words, keeping the initial phoneme and changing the remaining phonemes with similar sonority and manner of articulation; for example /*bava*/ (drool) became /*bofo*/, and /*galera*/ (prison) became /*gorelo*/. They also reported that they developed stimuli with low phonotactic probability. The current NWR stimuli can thus be considered language-like (Italian-like) non-words. Two NWR notation systems seem to be the most commonly used by the scientific community: word accuracy, which is the proportion of non-words that were correctly repeated (e.g., Chiat & Polišenská, 2016; Marini et al., 2017); and phoneme scoring, based on the proportion of phonemes correctly repeated (Munson, Kurtz & Windsor, 2005; Stokes, Moran & George, 2013). Moreover, different authors have tested the validity of the two scores and similar results have been found across scores overall (Dispaldro, Leonard & Deevy, 2013; Estes, Evans & Else-Quest, 2007) and in terms of correlations between NWR and other language measures (Brandeker & Thordardottir, 2015). In the current study, both scores were calculated: the proportion of items correctly repeated (whole-word scoring) and the proportion of phonemes correctly repeated (phoneme scoring). Concerning the whole-word scoring notation system, a correct

repetition was considered when all the phonemes were reported, in the same order, without any omissions and substitutions. Phoneme scoring was calculated as the number of vowels and consonants correctly repeated. Neither scoring penalized prosody or regional allophones. Only children's first repetition was collected. We also carried out inter-rater reliability on NWR scoring. All data was initially coded by the first author of the study and then blind-coded by an Italian cognitive science graduate student. Agreement assessment on NWR was substantial using whole-word scoring (Cohen's κ=.723), as well as using phoneme scoring (Cohen's κ=.698) (McHugh, 2012). Disagreements were solved by first author re-analysing the recordings of the NWR task. Since agreement was substantial, we decided to not analyze agreement on phoneme scoring further.

Stimuli were orally administered by the first author. Following the research design used by the authors who developed and published results of the stimuli we have used (Dispaldro, Leonard & Deevy, 2013), we decided to not audiorecord stimuli in order to create a more natural situation.

The Italian adaptation of the PPVT (Stella, Pizzoli & Tressoldi, 2000), a picture pointing test standardized on a population of monolingual Italian children, was used to estimate Italian receptive vocabulary size. The PPVT receptive vocabulary assessment was selected since it has been employed for different languages to assess lexical development. Participants have to choose among four pictures the one representing the word produced by the examiner. Two cut-offs define the beginning (basal) and the end (ceiling) of the test. The raw score, obtained by summing the number of pictures correctly identified, is transformed into a normalized score on the basis of the normative data (which is available for each age), giving a representative estimate of children's receptive vocabulary development.

The semi-structured interview about the linguistic background investigates sociolinguistic beliefs about bilingualism, immigration and migration history and plans,

information about the number of hours (out of 12 ~~a~~wake hours and separately for weekdays and weekend ~~day~~s) the child spent in daycare or kindergarten, with relatives or with others and the languages spoken in each of these settings, as well as language and psychomotor acquisition milestones and related concerns about linguistic development. It allows us to determine the onset and quantity of language input received in each language, as well as parental concerns on children's language development, as follows.

Cumulative language exposure is calculated as the sum of the time the child has been exposed to Italian. Parents reported the number of hours the child was exposed to Italian in a typical week, and this separately for relevant periods of the child's life. For example, if a child was cared at home until 18 months, and then at daycare up until the time of our assessment, then the typical weekday would be different in these two time periods. We assumed daycare and kindergarten holidays last eight weeks per year. Cumulative exposure is calculated up to MB-CDI administration, and up to NWR administration, in both hours and percentage of wake time, yielding four metrics of cumulative exposure. Current exposure is the amount of time the child has been exposed to Italian during the week before the NWR assessment, also expressed in hours and percent of ~~a~~wake time. Finally, length of exposure is calculated as the current age minus age of first exposure to Italian, expressed in months.

Parental concerns were estimated asking parents to judge, on a four-point scale, the level of satisfaction with their child's language proficiency in Italian. The four levels were: 1 "not at all satisfied" (*per niente soddisfatto*); 2 "satisfied enough" (*abbastanza soddisfatto*); 3 "satisfied" (*soddisfatto*); and 4 "completely satisfied" (*completamente soddisfatto*). Parental concerns were collected with the semi-structured interview questionnaire, which was administered to parents by the first author. All parents who filled the questionnaire spoke Italian and could complete the questionnaire in this language.

The proxy for socioeconomic status was maternal education level, collected in and extracted from the MB-CDI questionnaire; we did not ask for family income or other familiar background information, as this may not be culturally appropriate.

Statistical analysis

Since equal variances among groups cannot be assumed, several variables were transformed into percentages (e.g. NWR, cumulative and current exposure), one variable reported ranks (e.g. parental concerns), and a small sample size, we decided to carry out non-parametric analyses. We calculated Spearman Rho (r_s) correlations to study the relationship between the different variables, as well as to explore the predictive value of MB-CDI score in toddlerhood on PPVT and NWR in childhood (see Appendix D-F for alternative analyses). Given the relatively small sample size, we report direction and significance, as well as effect size, using the following cut-offs: Spearman Rho's coefficient of .1 is small, of .3 is moderate, and .5 and above is large (Cohen, 1981). Alpha level was set at alpha = .05 uncorrected; please see the Tables' captions for significance after Bonferroni correction for multiple comparisons.

Results

Individual descriptive data for each participant are provided in Table 1. The correlations between NWR accuracy (using both whole-word and phoneme scores) and estimated receptive vocabulary size were large, positive and statistically significant (Table 2). None of the language exposure measures was significantly correlated with the NWR accuracy measures, and their coefficients were small. NWR performance was not significantly correlated with maternal education, with a small effect size. Finally, the correlations between parental concerns and NWR scores were large, positive and significant: the less concerned the parents were about their child's language proficiency in Italian, the higher the NWR

accuracy in their child (Table 2). Please note that carrying out parametric analysis (see Appendix D), partialling out chronological age (see Appendix E), or considering non-words' lengths (Appendix F) does not alter our key results.

[Table 1 near here]

Receptive vocabulary scores significantly correlated with both age of Italian onset and cumulative exposure. This suggests our language exposure measures were sensitive to individual variation. Similarly to NWR scores, the receptive vocabulary score was significantly correlated with parental concerns: the lower children's score in receptive vocabulary was, the higher the index of parental concern about their child's Italian language proficiency.

[Table 2 near here]

Additionally, we explored the predictive value of MB-CDI scores measured in toddlerhood with PPVT and a NWR test, both collected in childhood. MB-CDI was not significantly correlated with receptive vocabulary scores nor NWR scores. In addition, children who had scored below the 10th percentile in MB-CDI as toddlers did not differ from the subgroup who had scored above that. Some readers may wonder whether this relates to the fact that there is variation in Italian exposure. To address this idea, we calculated the MB-CDI percentiles as a function of length of exposure to Italian rather than considering chronological age; e.g., for a 30-months-toddler who has been exposed to Italian for 24 months, the MB-CDI percentile will be derived by looking at 24 months normative data. Children whose length of exposure was lower than 18 months were scored using the 18 months' norms. The pattern of results obtained did not change: MB-CDI percentile derived from length of Italian exposure did not correlate with receptive vocabulary scores nor with NWR performance.

Discussion

The current study explored correlations between NWR performance and Italian receptive vocabulary size, as well as between NWR accuracy and language exposure, maternal education and parental concerns. Results contribute to ongoing debates on the association between NWR performance and both language experience and language development. We found significant correlations between NWR scores and both Italian estimated receptive vocabulary size and parental concerns about their child's Italian language development (both of large size, $r_s > .58$), and non-significant correlations between NWR accuracy and both language exposure and maternal education (both estimates near zero, $|r_s| > .02$). Additionally, we analyzed the predictive value of expressive vocabulary in toddlerhood with respect to childhood NWR performance ($r_s \sim .3$, of moderate size) and receptive vocabulary ($r_s \sim .1$, of small size). In the next paragraphs, we provide some possible interpretations of our data, and integrate it with evidence available in previous literature.

The relationship between NWR and receptive vocabulary, language exposure, maternal education and parental concerns

We found a positive, strong, and significant correlation between NWR performance and receptive vocabulary, in line with a growing body of literature in bilingual samples (e.g.: Core et al., 2017; Duncan & Paradis, 2016; White, 2020; but see also McBride-Chang et al., 2005; Stokes, Moran & George, 2013 and in particular Park & Schwarz, 2016 who found small to moderate non-significant association on 19 bilingual children). Our results are consistent with theoretical frameworks proposing that NWR and receptive vocabulary are associated in preschoolers because children use the same cognitive processes when learning words and when repeating non-words (e.g.: Marini et al., 2017), and/or because nonwords' lexical and sub-lexical cues enhance NWR performance directly (Coady & Evans, 2008).

Regarding language exposure, our results show small to moderate non-significant correlations between language exposure measures and NWR performance. Given the estimated sizes, it would be necessary to test over 100 participants to allow these correlations to become statistically significant; a larger sample size could clarify whether these correlations are non-significant due to the low number of participants. Alternatively, one hypothetical interpretation is that our measures of language exposure may not be appropriate. However, we did find larger correlations between the exact same language exposure measures and vocabulary size, suggesting that our measures of exposure were not flawed. We therefore repeated our inspection of previous literature on the relationship between language exposure and NWR performance. This actually revealed that while some work documents significant correlations (e.g., Engel de Abreu, 2011; Thordardottir, 2017), other work does not (e.g., Chiat & Polišenská, 2016; Core et al., 2017). One limitation of the current work is that we considered the quantity of exposure but not its quality. Quality of input might have a core role in explaining individual differences among bilingual children in their language-specific developmental paths. For example, factors such as exposure to native as opposed to non-native speakers, parental speech features, variability of speakers intra- and inter-languages (Carroll, 2016), as well as speakers' lexicon, syntactic complexity, and variability in the use of concrete as opposed to abstract conversations (Hoff et al., 2014b), all play an important role. Unfortunately, we did not have data to analyze such quality features. Another open avenue of research is to study whether NWR performance depends on total absolute quantity of input (cumulated across languages). Another limitation of this work is that we did not look at language use (i.e., children's use of the different languages in production) because it was not commonly collected in literature. However, it may be interesting to collect frequency of production as well as exposure in the future. Finally, another possible interpretation is that of a true null result: perhaps performance in a

language-like NWR task like the one we used is not related to language exposure and input (Cristia et al., 2020). Further research is necessary to address all of these possibilities.

Regarding our remaining factors, we found that maternal education was not significantly related to NWR performance, with an effect size very close to zero. This result is in line with interpretations proposed by previous work on the relationship between SES and NWR performance (Chiat & Polišenská, 2016; Meir & Armon-Lotem, 2017). We also highlight that maternal education on its own does not capture SES completely, so further studies could revisit this question with alternative proxies of SES.

Looking at parental concerns, we found a significant and large correlation with children's NWR accuracy: the more concerned parents were about their child language development, the lower the child's NWR performance. We pointed out in the Introduction that few articles have analyzed the association between non-word repetition performance and parental concerns. In line with a recent study (de Almeida et al., 2017), our results suggest that both non-word repetition performance and parental concerns are reliable proxies of children's language development. Moreover, parents' and children's psychosocial risk factors should be assessed to study their association with language development (St Clair et al., 2019). Further studies on multifactorial language assessment in bilingual children should explore accuracy in identifying atypical language development paths, including variables such as NWR and parental concerns.

As briefly mentioned in the Introduction, results may differ for different NWR lists, depending on their language-specific properties. We followed previous work by using one language-specific NWR list (e.g., Guiberson & Rodríguez, 2015), which may be viewed as Italian-like. Further articles should study the effect of lexical cues on NWR performance, as well as its associations with both children's and environmental factors, in order to identify the optimal NWR research design to study bilingual language assessment for research and

clinical purposes. Finally, it could be interesting to build a crosslinguistic NWR assessment tool for Italian-speaking bilingual children and compare data with our results on Italian-like NWR, in order to explore different NWR research designs and improve the current bilingual language assessment procedures.

NWR performance in Italian-speaking bilinguals as a function of toddler vocabulary in the majority language

The sample that we studied had received a previous lexical assessment in toddlerhood in the major language, i.e., Italian. Even though our sample is not representative of bilingual populations everywhere, we underline that we would like to contribute to the study of predictive values of language acquisition paths across time (Rice, 2016). We thus explored whether the score in a monolingual assessment (Italian MB-CDI) in toddlerhood was related with later linguistic development in childhood.

Our results showed a non-significant but moderate correlation between expressive vocabulary (MB-CDI) in toddlerhood and our NWR task, as well as a non-significant and small correlation between MB-CDI and receptive vocabulary (PPVT) in childhood. Different possible interpretations can be posited for these results. One of the main interpretations is that some of these children had language difficulties at toddlerhood, but by childhood they filled their Italian language-specific linguistic gap (i.e., they were "late bloomers"). Another possible interpretation is that the assessment they received in toddlerhood only in the major language was unreliable for different reasons. First, the Italian MB-CDI was normed on monolingual children, while our sample is composed by bilingual children exclusively. It has been argued that a language-specific language assessment normed on a monolingual sample has a low predictive value of bilingual language development (Haman, Łuniewska & Pomiechowska, 2015). Second, perhaps the short-form version of MB-CDI is more of a screening tool than an assessment to identify language development difficulties.

It is relevant to compare our results to the ones obtained with the same NWR stimuli on a sample of monolingual Italian-speaking children (Dispaldro, Leonard & Deevy, 2013). Although we used the same stimuli, there are several differences between the two studies. To begin with, the current study involved bilingual children, while the other study concerned monolingual children. In addition, they recruited children with and without a diagnosis of DLD at the time of testing. Overall, we found lower NWR scores on bilinguals compared to monolingual performance: 51% whole-word and 87% phoneme scoring in the current study compared to 83% and 97% respectively for monolingual children with typical language development in Dispaldro et al. (2013). This is in line with previous work reporting lower scores for bilinguals than monolinguals in language-specific NWR (Armon-Lotem & Meir, 2016; Kohnert, Windsor & Yim, 2006; Gutiérrez-Clellen & Simon-Cereijido, 2010; Messer et al., 2010).

Clinical implications

In this section, we would like to weigh in on the ongoing debate regarding bilingual language assessment in childhood, as well as discuss some realistic solutions available to clinicians who see multilingual children in their practice, in geographical areas where there are multilingual populations with different native languages. In a nutshell, we recommend a culturally sensitive evaluation, which makes informed use of assessments in all of the child's languages (when such assessments are available), keeping in mind the subjectivity and specificity of each child.

One set of arguments in the debate runs as follows. Clinicians should analyze language development in all the languages the child has experienced, at least in the form of qualitative analysis. Even when speech and language therapists are not proficient in all the languages the child is exposed to, they can administer communicative and linguistic parental

questionnaires in either bilingual or monolingual versions of the target languages. We stress that quantitative and qualitative language exposure features should be analyzed during a language assessment including observations of parent-child interactions and parental questionnaires.

Given our results suggesting that MB-CDI was not a good index of later language, one can argue that bilingual language assessment tools normed on bilingual samples should be preferred over tools developed on monolinguals only, and, considering our discussions regarding crosslinguistic NWR stimuli, crosslinguistic tasks could also be favored. For children younger than 3, parental questionnaires can be used to analyze lexical development and related features, using versions corresponding to each language the child is exposed to during their lifespan. In those cases where clinicians do not have access to normed parental questionnaires of a target native language, parents could complete a questionnaire edited in a language they are proficient in but thinking about their children's language development in the other native language.

Further work should study the optimal lexical measures to analyze bilingual lexical development, since mixed evidence has been found using total vocabulary (Core et al., 2013), which is the total number of words known by a child in all the languages she is exposed to, or conceptual vocabulary (Bedore et al., 2005), that is the total number of concepts known across languages.

Moreover, clinicians can use their clinical and common sense to interpret these data: for example, reduced exposure to a target language should induce clinicians to give more importance to results achieved in the primary language over those achieved in languages to which the child has been less exposed.

While the recommended approach laid out above is ideal, socio-cultural and sociolinguistic issues may not render it possible in all situations. For example, when the child

speaks three or more languages, when there is no parental questionnaire for some of these languages, when the native language is not written or the parent does not read/write in all of these languages, and/or when the parent does not use all languages with the child. For all these cases where bilingual language assessment tools are not available, monolingual norms of monolingual normed tests in the languages spoken by the child, integrated with qualitative considerations that are culturally sensitive, appear like the optimal solution, avoiding the claim that the label "bilingual" is a unitary identity, when in fact it applies to a heterogeneous population. Additionally, dynamic assessments could also be considered to study bilinguals' language development (Petersen et al., 2017).

We also wondered about the utility of a complementary approach, namely combining language exposure features with lexical development in only the main language. This solution appears feasible even in geographical areas where bilinguals have diverse native languages and they share a single major language. It would involve administering a test in the major language as well as a short and simple language exposure questionnaire, which simply asks what proportion of an average week is spent hearing different languages. The clinician can then qualitatively or mathematically integrate a consideration of the proportion of time spent listening to the main language as well as other factors describing the child (e.g., migration features, siblings, family history of language disorders, risk factors of language disorders) and the child's input (e.g., where, from who and from how many people languages are spoken, lexical and syntax richness of language exposure).

Conclusions

We find that language-like NWR scores are strongly related to receptive vocabulary scores and parental concerns, with non-significant and weaker links to language exposure and maternal education. Our data are consistent with the hypothesis that NWR reflects language

proficiency beyond the effect of lingual status and socioeconomic status.

The use of NWR as a diagnostic tool to identify atypical language developmental paths is still a debated topic in the scientific community. On the one hand, since in the literature bilingual NWR lists are useful when identifying bilingual children with DLD, it might be useful for both clinical and research purposes to create sensitive and specific NWR lists constructed in such a way that they are in accordance with various constraints of phonological systems of different languages (Armon-Lotem, de Jong & Meir, 2015). On the other hand, in heterogeneous bilingual populations with different native languages, it is challenging to develop crosslinguistic stimuli that consider phonological constraints of all the native languages spoken in a given geographical area. We hope these considerations are taken into account in the future of both research and practice.

To the best of our knowledge, this is the first study on NWR performance in a sample of Italian-speaking bilingual children exploring the predictive value of expressive vocabulary in the major language at toddlerhood. Results suggested that MB-CDI scores in toddlerhood are not significantly related with NWR accuracy (small to moderate correlations) or with receptive vocabulary scores in childhood (small correlations). In conclusion, despite the fact that assessing language development in bilingual children with a vocabulary checklist in the major language is the most feasible procedure in the majority of healthcare services, our results suggest that monolingual MB-CDI at toddlerhood and NWR in early childhood may not be as informative for bilinguals as they are for monolinguals.

Statements

Statement of Ethics

The research was conducted ethically in accordance with the World Medical Association Declaration of Helsinki. Participants' parents have given their written informed consent and that the study protocol was approved by the institute's committee on human research.

Disclosure of interest

The authors have no conflicts of interest to declare.

Funding Sources

Authors acknowledge the support of the National Institute for Health, Migration and Poverty (INTEGRA: Identificazione e intervento su bambini a rischio di disturbi del linguaggio e/o del neuro-sviluppo, figli di migranti); the Agence Nationale de la Recherche (ANR-17-CE28-0007 LangAge, ANR-16-DATA-0004, ANR-14-CE30-0003, ANR-17-EURE-0017); and the J. S. McDonnell Foundation (Understanding Human Cognition Scholar Award).

Data availability statement.

We present descriptive statistics with data collected on our sample.

Author Contributions

The second author contributed to statistical analysis, in the data interpretation and in writing the manuscript. The third author contributes to the writing of the manuscript, with an additional role of senior researcher. The last author contributed to elaborating the research design, the analysis and interpretation of data and in the writing of the manuscript. The first author contributed to building the research design, in the data interpretation and in the writing; he has also carried out both the review of the literature and data collection.

References

Abed Ibrahim, L., & Fekete, I. (2019). What Machine Learning Can Tell Us About the Role of Language Dominance in the Diagnostic Accuracy of German LITMUS Non-word and Sentence Repetition Tasks. *Frontiers in psychology*, *9*, 2757.

Altman, C., Burstein Feldman, Z., Yitzhaki, D., Armon Lotem, S., & Walters, J. (2014). Family language policies, reported language use and proficiency in Russian–Hebrew bilingual children in Israel. *Journal of Multilingual and Multicultural Development*, *35*(3), 216-234.

American Speech-Language-Hearing Association (ASHA) (2004). Knowledge and skills needed by speech-language pathologists and audiologists to provide culturally and linguistically appropriate services. Retrieved from: https://www.asha.org/policy/KS2004-00215.htm.

Armon-Lotem, S., de Jong, J., & Meir, N. (2015). *Assessing multilingual children: Disentangling bilingualism from language impairment*. Bristol, UK: Multilingual Matters. https://doi.org/10.21832/9781783093137-009.

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(6), 715–31. https://doi.org/10.1111/1460-6984.12242.

Baddeley, A. (1986). *Working Memory*. Oxford, UK: Clarendon Press.

Balladares, J., Marshall, C., & Griffiths, Y. (2016). Socio-economic status affects sentence repetition, but not non-word repetition, in Chilean preschoolers. *First Language*, 36, 338–351. https://doi.org/10.1177%2F0142723715626067.

Bedore, L. M., Peña, E. D., Joyner, D., & Macken, C. (2011). Parent and teacher rating of bilingual language proficiency and language development concerns. *International Journal of Bilingual Education and Bilingualism*, 14(5), 489-511. https://dx.doi.org/10.1080%2F13670050.2010.529102.

Bishop, D. V. M., Snowling, M. J., Thompson, P. A., & Greenhalgh, T. (2017). The Catalise Consortium: CATALISE: a multinational and multidisciplinary Delphi consensus study of problems with language development. Phase 2. Terminology. *Journal of Child Psychology and Psychiatry*, 58, 1068–1080. https://doi.org/10.1111/jcpp.12721.

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), 1747–60. https://doi.org/10.1044/2015_JSLHR-L-15-0058.

Brandeker, M., & Thordardottir, E. (2015) Language exposure in bilingual toddlers: Performance on nonword repetition and lexical tasks. *American Journal of Speech-Language Pathology*, 24(2), 126-138. https://doi.org/10.1044/2015_ajslp-13-0106.

Carroll, S. E. (2016). Explaining bilingual learning outcomes in terms of exposure and input. *Bilingualism: Language and cognition*, 20(1), 37–41. https://doi.org/10.1017/S1366728916000511.

Casalini, C., Brizzolara, D., Chilosi, A., Cipriani, P., Marcolini, S., Pecini, C., & Burani (2007). Non-word repetition in children with specific language impairment: A deficit in phonological working memory or in long-term verbal knowledge? *Cortex*, 43(6), 769-776.

Chiat, S., & Polišenská, K. (2016). A Framework for crosslinguistic nonword repetition tests: effects of bilingualism and socioeconomic status on children's performance. *Journal of Speech, Language, and Hearing Research*, 59(5), 1179–89. https://doi.org/10.1044/2016_JSLHR-L-15-0293.

Chilosi, A. M., Pfanner, L., Pecini, C., Salvadorini, R., Casalini, C., Brizzolara, D., & Cipriani, P. (2019). Which linguistic measures distinguish transient from persistent language problems in Late Talkers from 2 to 4 years? A study on Italian speaking children. *Research in Developmental Disabilities*, 89, 59-68. https://doi.org/10.1016/j.ridd.2019.03.005.

Coady, J., & Evans, J. L. (2008). Uses and interpretations of non-word repetition tasks in children with and without specific language impairments (SLI): *International Journal of Language & Communication Disorders*, 43(1), 1–40. https://dx.doi.org/10.1080%2F13682820601116485.

Coady, J., Evans, J. L. & Kluender, K. R., (2010): Role of phonotactic frequency in nonword repetition by children with specific language impairments. *International journal of language & communication disorders / Royal College of Speech & Language Therapists*, 45(4), 494–509. https://doi.org/10.3109/13682820903222783.

Cohen, P. A. (1981). Student ratings of instruction and student achievement: A meta-analysis of multisection validity studies. *Review of educational Research*, 51(3), 281-309. https://doi.org/10.3102%2F00346543051003281.

Core, C., Chaturvedy, S., & Martinez-Nadramia, D. (2017). The role of language experience in nonword repetition tasks in young bilingual Spanish-English speaking children. *Proceedings of the 41st annual Boston University Conference on Language Development*. Boston, US: Cascadilla Press; 2017. Available from: http://www.lingref.com/bucld/41/BUCLD41-14.pdf.

Core, C., Hoff, E., Rumiche, R., & Señor, M. (2013). Total and conceptual vocabulary in spanish-english bilinguals from 22 to 30 months: Implications for assessment. *Journal of Speech, Language, and Hearing Research*, 56, 1637–49. https://doi.org/10.1044/1092-4388(2013/11-0044).

Cristia, A., Farabolini, G., Scaff, C., Havron, N., & Stieglitz, J. (2020). Infant-directed input and literacy effects on phonology: Non-word repetition accuracy among the Tsimane. *PlosONE*, 15(9), e0237702. https://doi.org/10.1371/journal.pone.0237702.

D'odorico, L., Assanelli, A., Franco, F., & Jacob, V. (2007). A follow-up study on Italian late talkers: Development of language, short-term memory, phonological awareness, impulsiveness, and attention. *Applied Psycholinguistics*, 28(1), 157–69. https://doi.org/10.1017/S0142716406070081.

de Almeida, L., Ferre, S., Morin, E., Prevost, P., Dos Santos, C., Tuller, L., Zebib, R., & Barthez, M. (2017). Identification of bilingual children with Specific Language

Impairment in France. *Linguistic Approaches to Bilingualism*, 7(3), 331-358. https://doi.org/10.1075/lab.15019.alm.

Desmarais, .C, Sylvestre, A., Meyer, F., Bairati, I., & Rouleau., N. (2008). Systematic review of the literature on characteristics of late‐talking toddlers. *International Journal of Language & Communication Disorders*, 43(4), 361–89. https://doi.org/10.1080/13682820701546854.

Dispaldro, M., Leonard, L. B., & 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), 323–36. https://dx.doi.org/10.1044%2F1092-4388(2012%2F11-0304).

Duncan, T. S., & Paradis, J. (2016). English language learners' nonword repetition performance: The Influence of age, L2 vocabulary size, length of L2 exposure, and L1 phonology. *Journal of Speech, Language, and Hearing Research*, 59(1), 39–48. https://doi.org/10.1044/2015_JSLHR-L-14-0020.

Edwards J., Beckman M. E., Munson B. (2004). The interaction between vocabulary size and phonotactic probability effects on children's production accuracy and fluency in nonword repetition. *Journal of Speech, Language, and Hearing Research,* 47(2), 421–36. https://doi.org/10.1044/1092-4388(2004/034).

Engel de Abreu, P. M. J., Baldassi, M., Puglisi, M. L., & Befi-Lopes, D. M. (2013). Cross-linguistic and cross-cultural effects on verbal working memory and vocabulary: Testing language-minority children with an immigrant background. *Journal of Speech, Language, and Hearing Research*, 56(2), 630– 642. https://doi.org/10.1044/1092-4388(2012/12-0079).

Engel de Abreu, P. M. J. (2011). Working memory in multilingual children: Is there a bilingual effect? *Memory*, 19(5), 529–37. https://doi.org/10.1080/09658211.2011.590504.

Estes, G. K., Evans, J. L., & Else-Quest, N. M. (2007). Differences in the nonword repetition performance of children with and without Specific Language Impairment: A

Meta-Analysis. *Journal of Speech, Language, and Hearing Research*, 50(1), 177–95. https://doi.org/10.1044/1092-4388(2007/015).

Fenson, L., Marchman, V.A., Thal, D.J., Dale, P.S., Reznick, J.S., Bates, E. (2007). *MacArthur-bates communicative development inventories (CDIs). User's guide and technical manual*. Baltimore, US: Brookes Publishing (second edition).

Frank, M. C., Braginsky, M., Yurovsky, D., & Marchman, V. A. (2017). Wordbank: an open repository for developmental vocabulary data. *Journal of Child Language*, 44(3), 677–94. https://doi.org/10.1017/S0305000916000209.

Gagarina, N., Bohnacker, U., & Lindgren, J. (2019). Macrostructural organization of adults' oral narrative texts. Narrative texts by children and adults: insights into their organization through a prism of language system and environmental factors. *Leibniz-Zentrum Allgemeine Sprachwissenschaft*, 190. Available from: https://www.leibniz-zas.de/en/research/publications/details/publications/4386-narrative-texts-by-children-and-adu/.

Gagarina, N., Klop D., Kunnari S., Tantele K., Välimaa T., Balčiūnienė I., ... & Walters J. (2012). MAIN: Multilingual assessment instrument for narratives. *Zentrum für Allgemeine Sprachwissenschaft*. Available from: https://www.leibniz-zas.de/publications/schriftenreihe/zaspil/zaspil-56/main-start/.

Gatt, D., & O'Toole, C. (2016). Risk and protective environmental factors for early bilingual language acquisition. *International Journal of Bilingual Education and Bilingualism*, 20(2), 117–23. https://doi.org/10.1080/13670050.2016.1179926.

Gatt, D., O'Toole, C., & Haman, E. (2015). Using Parental Report to Assess Early Lexical Production in Children Exposed to More than One. In *Assessing multilingual children: Disentangling bilingualism from language impairment* (eds. Armon-Lotem. de Jong & Meir, 2015), 13, 151. https://doi.org/10.21832/9781783093137-009.

Gibson, T. A., Summers, C, Peña, E. D., Bedore, L. M., Gillam, R. B., & Bohman, T. M. (2015). The role of phonological structure and experience in bilingual children's nonword repetition performance. *Bilingualism: Language and Cognition*, 18(3), 551–60. https://doi.org/10.3109/13682820903198058.

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(2), 124–45. https://dx.doi.org/10.1016%2Fj.jcomdis.2007.07.001.

Grimm, A., Ferré, S., dos Santos, C., & Chiat, S. (2014). Can nonwords be language-independent? Cross-linguistic evidence from monolingual and bilingual acquisition of French, German, and Lebanese. In *Symposium Language Impairment Testing in Multilingual Setting (LITMUS): Disentangling bilingualism and SLI*.

Guiberson, M., & Rodríguez, B.L. (2015). Classification accuracy of nonword repetition when used with preschool-age Spanish-speaking children. *Language, Speech, and Hearing Services in Schools*, 67(5), 253-258. https://doi.org/10.1159/000442745.

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(1), 48–58. https://doi.org/10.1111/j.1540-5826.2009.00300.x.

Haman, E., Wodniecka, Z., Marecka, M., Szewczyk, J., Białecka-Pikul, M., Otwinowska, A., Mieszkowska, K., Łuniewska, M., Kołak, J., Miekisz, A., Kacprzak, A., Banasik, N. & Forys-Nogala, M. (2017). How does L1 and L2 exposure impact L1 performance in bilingual children? Evidence from Polish-English migrants to the United Kingdom. *Frontiers in psychology*, 8,1444. https://doi.org/10.3389/fpsyg.2017.01444.

Haman, E., Łuniewska, M., & Pomiechowska, B. (2015). Designing cross-linguistic lexical tasks (CLTs) for bilingual preschool children. In *Assessing multilingual children: Disentangling bilingualism from language impairment* (eds. Armon-Lotem, de Jong & Meir, 2015), 196-240. https://doi.org/10.21832/9781783093137-009.

Hoff, E., Core, C., Place., S., Rumiche, R., Señor, M., & Parra, M. (2012). Dual language exposure and early bilingual development. *Journal of Child Language*, 39(1), 1-27. https://doi.org/10.1017/S0305000910000759.

Hoff, E., Rumiche, R., Burridge, A., Ribot, K. M., & Welsh, S. N. (2014)a. Expressive vocabulary development in children from bilingual and monolingual homes: A

longitudinal study from two to four years. *Early childhood research quarterly*, 29(4), 433-444. https://doi.org/10.1016/j.ecresq.2014.04.012.

Hoff, E., Welsh, S., Place, S., Ribot, K., Grüter, T., & Paradis, J. (2014)b. Properties of dual language input that shape bilingual development and properties of environments that shape dual language input. *Input and experience in bilingual development*, 13, 119-140. https://doi.org/10.1075/tilar.13.07hof.

Hoff, E. (2020). Lessons from the study of input effects on bilingual development. *International Journal of Bilingualism*, 24(1), 82-88. https://doi.org/10.1177%2F1367006918768370.

Huls, S.K. (2017). Bilingual children's performance on three nonword repetition tasks: the role of language experience and ability (Doctoral dissertation). US: University of Kansas. Retrieved from: https://kuscholarworks.ku.edu/bitstream/handle/1808/26333/Huls_ku_0099D_15325_DATA_1.pdf?sequence=1&isAllowed=y.

International Association of Logopedics and Phoniatrics (IALP) (2011). Recommendations for working with bilingual children. Retrieved from: http://www.specchioriflesso.net/media/162083/linee_guida_bilingui_ialp-_may_2011.pdf.

Kohnert, K., Windsor, J., & Yim, D. (2006). Do language-based processing tasks separate children with language impairment from typical bilinguals?. *Learning Disabilities Research & Practice*, 21(1), 19–29. https://doi.org/10.1111/j.1540-5826.2006.00204.x.

Lee, H. J., Kim, Y. T., & Yim, D. (2013). Non-word repetition performance in Korean-English bilingual children. *International Journal of Speech-Language Pathology*, 15(4), 375–82. https://doi.org/10.3109/17549507.2012.752866.

Lee, S. A. S.,& Gorman, B. K. (2013). Nonword repetition performance and related factors in children representing four linguistic groups. *International Journal of Bilingualism*, 17(4), 479-495. https://doi.org/10.1177%2F1367006912438303.

Li'el, N. (2017). Assessing dual language learners with primary language impairments: Developing an assessment approach (Doctoral dissertation). US: Curtin University. https://doi.org/10.1177%2F1367006912438303.

Marini, A., Ruffino, M., Sali, M. E., & Molteni, M. (2017). The role of phonological working memory and environmental factors in lexical development in Italian-speaking late talkers: A one-year follow-up study. *Journal of Speech, Language, and Hearing Research*, 60(12), 3462–73. https://doi.org/10.1044/2017_JSLHR-L-15-0415.

McBride-Chang, C., Wagner, R. K., Muse, A., Chow, B. W. Y., & Shu, H. U. A. (2005). The role of morphological awareness in children's vocabulary acquisition in English. *Applied psycholinguistics*, 26(3), 415-435. https://doi.org/10.1017/S014271640505023X.

McHugh, M. L. (2012). Interrater reliability: the kappa statistic. *Biochemia medica: Biochemia medica*, 22(3), 276-282. PMID: 23092060.

Meir, N., & Armon-Lotem, S. (2017). Independent and combined effects of socioeconomic status (SES) and bilingualism on children's vocabulary and verbal short-term memory. *Frontiers in Psychology*, 8, 1442. https://dx.doi.org/10.3389%2Ffpsyg.2017.01442.

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(4), 306–23. https://doi.org/10.1016/j.jecp.2009.12.006.

Munson, B., Kurtz, B. A., & Windsor J. (2005). The influence of vocabulary size, phonotactic probability, and wordlikeness on nonword repetitions of children with and without specific language impairment. *Journal of Speech, Language, and Hearing Research*, 48(5), 1033–47. https://doi.org/10.1044/1092-4388(2005/072).

Onofrio, D., Rinaldi, P., & Pettenati, P. (2012). Il primo sviluppo del linguaggio in bambini che imparano più lingue: una proposta per la valutazione e l'interpretazione del

profilo linguistico. *Psicologia clinica dello sviluppo*, *16*(3), 661-670. doi: 10.1449/38845.

Onofrio, D. (2017). Riconoscere un ritardo di linguaggio nei bambini bilingui. Studio delle traiettorie evolutive e degli indicatori precoci di rischio nello sviluppo del linguaggio. (PhD dissertation). Rome, Italy: Università degli Studi di Roma La Sapienza.

Park, H., & Schwarz, I. (2016). The influence of language proficiency on language-based processing skills: evidence from Korean–English bilingual children. *Speech, Language and Hearing*, 19(4), 189–202. https://doi.org/10.1080/2050571X.2016.1169623.

Parra M., Hoff E., & Core C. (2011). Relations among language exposure, phonological memory, and language development in Spanish–English bilingually developing 2-year-olds. *Journal of Experimental Child Psychology*, 108(1), 113–25. https://dx.doi.org/10.1016%2Fj.jecp.2010.07.011.

Pearson, B. Z., Fernández, S., & Oller, D. K. (1993). Lexical development in bilingual infants and toddlers: Comparison to monolingual norms. *Language Learning.* 43, 93–120. https://doi.org/10.1111/j.1467-1770.1993.tb00174.x.

Petersen, D. B., Chanthongthip, H., Ukrainetz, T. A., Spencer, T. D., & Steeve, R. W. (2017). Dynamic assessment of narratives: Efficient, accurate identification of language impairment in bilingual students. Journal of Speech, Language, and Hearing Research, 60(4), 983-998. https://doi.org/10.1044/2016_JSLHR-L-15-0426.

Petruccelli, N., Bavin, E. L., Bretherton, L. (2012). Children with specific language impairment and resolved late talkers: Working memory profiles at 5 years. *Journal of Speech, Language, and Hearing Research*. 55(6), 1690–1703. https://doi.org/10.1044/1092-4388(2012/11-0288).

Place, S., & Hoff, E. (2011). Properties of dual language exposure that influence two-year-olds' bilingual proficiency. *Child Development*, 82(6), 1834–49. https://dx.doi.org/10.1016%2Fj.jecp.2010.07.011.

Reilly, S., Cook, F., Bavin, E.L., Bretherton, L., Cahir, P., Eadie, P., Gold, L., … Wake, M. (2018). Cohort profile: The early language in Victoria student (ELVS). *International Journal of Epidemiology*, 47(1), 11-20. https://doi.org/10.1093/ije/dyx079.

Rescorla, L. (1989). The language development survey: a screening tool for delayed language in toddlers. *Journal of Speech and Hearing Disorders*, 54(4), 587–99. https://doi.org/10.1044/jshd.5404.587.

Rice, M. L. (2016). Specific language impairment, nonverbal IQ, attention-deficit/hyperactivity disorder, autism spectrum disorder, cochlear implants, bilingualism, and dialectal variants: Defining the boundaries, clarifying clinical conditions, and sorting out causes. *Journal of Speech, Language, and Hearing Research*, 59(1), 122-132. https://doi.org/10.1044/2015_JSLHR-L-15-0255.

Rinaldi, P., Pasqualetti, P., Stefanini, S., Bello, A., & Caselli, M. C. (2019). The Italian Words and Sentences MB-CDI: normative data and concordance between complete and short forms. Journal of child language, 46(3), 546-566. https://doi.org/10.1017/S0305000919000011.

Rispens, J., Baker, A., & Duinmeijer, I. (2015). Word recognition and nonword repetition in children with language disorders: the effects of neighborhood density, lexical frequency, and phonotactic probability. *Journal of Speech, Language, and Hearing Research*, 58(1), 78–92. https://doi.org/10.1044/2014_JSLHR-L-12-0393

Saiegh-Haddad, E., & Ghawi-Dakwar, O. (2017). Impact of diglossia on word and non-word repetition among language impaired and typically developing Arabic native speaking children. *Frontiers in Psychology*, 8, 2010. https://doi.org/10.3389/fpsyg.2017.02010.

Santos Dos, C., & Ferré, S. (2016). A nonword repetition task to assess bilingual children's phonology. *Language Acquisition*, 25(1), 58-71. https://doi.org/10.1080/10489223.2016.1243692.

Sharp, K. M., & Gathercole, V. (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(1), 77–89. https://doi.org/10.1177%2F0265659012465208.

St Clair, M. C., Forrest, C. L., Yew, S. G. K., & Gibson, J. L. (2019). Early risk factors and emotional difficulties in children at risk of developmental language disorder: A population cohort study. *Journal of Speech, Language, and Hearing Research*, 62(8), 2750-2771. https://doi.org/10.1044/2018_JSLHR-L-18-0061.

Stella, G., Pizzoli, C., & Tressoldi, P. (2000). *PPVT-R- Peabody Picture Vocabulary Test-R. Italian adaptation and standardization*. Torino, Italy: Omega Edizioni.

Stokes, S.F., Moran, C., & George, A. (2013). Nonword repetition and vocabulary use in toddlers. *Topics in Language Disorders*, 33(3), 224–237. https://psycnet.apa.org/doi/10.1097/TLD.0b013e31829d038c .

Summers, C., Bohman, T. M., Gillam, R. B., Peña, E. D., & Bedore, L. M. (2010). Bilingual performance on nonword repetition in Spanish and English. *International Journal of Language & Communication Disorders*, 45(4), 480–93. https://doi.org/10.3109/13682820903198058.

Thordardottir, E., & Brandeker, M. (2013). The effect of bilingual exposure versus language impairment on nonword repetition and sentence imitation scores. *Journal of Communication Disorders*, 46(1), 1–16. https://doi.org/10.1016/j.jcomdis.2012.08.002.

Thordardottir, E., & Juliusdottir, A. G. (2013). Icelandic as a second language: A longitudinal study of language knowledge and processing by school-age children. *International Journal of Bilingual Education and Bilingualism*, 16(4), 411–35. https://doi.org/10.1080/13670050.2012.693062.

Thordardottir, E. (2017). Are background variables good predictors of need for L2 assistance in school? Effects of age, L1, amount, and timing of exposure on Icelandic language and nonword repetition scores. *International Journal of Bilingual Education and Bilingualism*, 23(4), 1–23. https://doi.org/10.1080/13670050.2017.1358695.

Tuller, L., Hamann, C., Chilla, S., Ferré, S., Morin, E., Prevost, P., ... , & Zebib, R. (2018). Identifying language impairment in bilingual children in France and in Germany.

International journal of language & communication disorders, 53(4), 888-904. https://doi.org/10.1111/1460-6984.12397.

Vender, M., Delfitto, D., & Melloni, C. (2018). Clitic production in bilingual children: when exposure matters. *Languages*, 3(3), 22. https://doi.org/10.3390/languages3030022.

Verhoeven, L., Steenge, J., & van Balkom, H. (2012). Linguistic transfer in bilingual children with specific language impairment. *International Journal of Language & Communication Disorders*, 47(2), 176-183. https://psycnet.apa.org/doi/10.1111/j.1460-6984.2011.00092.x.

White, M. J. (2020). Phonological working memory and non-verbal complex working memory as predictors of future English outcomes in young ELLs. *International Journal of Bilingualism*. https://doi.org/10.1177/1367006920948136.

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. https://doi.org/10.1044/1058-0360(2010/09-0064).

Table and Figure Legends

Table 1: Descriptive data of our bilingual sample.

Table 2: Correlation matrix among studied variables.

Table A: Correlation matrix considering nonwords' word length.

Appendix A: Previous work's correlation results between NWR performance and lexical scores on bilingual children.

Study	N	Age (months)	A/Typical sample	NWR tool	Lexical assessment	Correlation	p value
Core et al. (2017)	105	22	typical	Bilingual Spanish- and English-like	Bilingual expressive vocabulary (Spanish and English MB-CDI)	.43	.000
						.27	.005
Duncan & Paradis (2016)	75	69,6	typical	Monolingual English-like nonwords	English estimated receptive vocabulary size (PPVT)	β= .010, z= 2.017	.044
Girbau et al. (2008)	22[1]	103,2	atypical	Spanish-like nonwords	Spanish lexical fluency task	.57	< .01
Guiberson & Rodriguez (2015)	54[2]	35.5	mixed	Spanish-like nonwords	Expressive (Spanish MB-CDI)	.52	<0.001
					Combined Spanish expressive and receptive vocabulary	.44	<0.001
Lee, Kim & Yim (2013)	30	54,3	typical	Monolingual Korean-like nonwords	English and Korean estimated receptive vocabulary size (PPVT)	.381	< .05
Messer et al. (2010)	60	52.5	typical	Turkish- and Dutch-like NWR (low PP)	Bilingual test for Turkish-Dutch bilinguals	ß = -0.23	>0.05
				Turkish- and Dutch-like NWR (high PP)		ß = 0.10	>0.05
				Turkish- and Dutch-like NWR (low PP)		ß = 0.25	>0.05
				Turkish- and Dutch-like NWR (high PP)		ß = 0.31	<0.05
Park & Schwarz (2016)	19	96	typical	Bilingual Korean- and English-like nonwords	Bilingual receptive vocabulary	.261	>.05
						.014	>.05

[1] children with typical development (N = 11) and with specific language impairment (N = 11)

[2] children with typical (N = 23) and (N = 21) atypical language development

Appendix B: Previous work's correlation results between NWR performance and language exposure measures on bilingual children

Study	N	Age (months)	NWR type	Language exposure measure	Correlation	Significance
Chiat & Polišenská (2016)	42[3]	66,9	Crosslinguistic	monolingual vs bilingual NWR performance	$F(1, 37) = 0.17$, $\eta_p^2 = .00$	.682
			English prosodically-like		$F(1, 37) = 1.07$, $\eta_p^2 = .03$	.307
			English-like		$F(1, 37) = 0.84$, $\eta_p^2 = .02$	.366
Core et al. (2017)	135	22	Spanish-like	Current Spanish exposure	.17	.08
			English-like		.05	.63
de Almeida et al. (2017)	82[4]	75.35	Crosslinguistic	age of contact	-0.02	0.866
				length of exposure	0.107	0.411
				use of French at home	0.158	0.222
				use of French during activities and with friends	0.09	0.489
				language dominance index	0.133	0.306
				age of contact	0.074	0.766
				length of exposure	0.091	0.701
				use of French at home	0.294	0.208
				use of French during activities and with friends	0.152	0.521
				language dominance index	0.312	0.181
Duncan & Paradis (2016)	75	68	English-like	Cumulative exposure	$\beta = 0.040$, $z = 2.548$	.011

[3] 21 monolinguals and 21 bilinguals

[4] 61 monolinguals and 21 bilinguals

Engel de Abreu (2011)	44[5]	76,8	Luxembourgish-like	monolingual vs bilingual NWR performance	F(1, 42) = 5.21, η_p^2 = .11	< .05
Gibson et al. (2015)	52	69,9	Spanish- and English-like	Current and cumulative exposure	F(1,50) = 5.07, η_p^2 = .09)	.03
Huls (2017)	47	71	Crosslinguisitic	Age of first exposure to English	F(5, 41)= 1.82	.13
			English-like		F(5, 41)= 2.00	.1
			Spanish- and English-like		F(5, 41)= 3.09	.02
Kohnert et al. (2006)	72[6]	125,46	English-like	monolingual vs bilingual NWR performance	F(1,69) = 7.71, d = 0.84	.007
Lee & Gorman (2012)	57[7]	84	English-like	monolingual vs bilingual NWR performance	F (3, 53) = 2.28, η = 0.11	.09
Li'el (2017)	61	70.14	English-like	Age of first exposure to English	0.275	< .05
				Length of Exposure	-0.222	< .05
		68.58[8]		Age of first exposure to English	0.098	< .05
				Length of Exposure	-0.143	< .05
Parra, Hoff & Core (2011)	41	30	English-like	Cumulative Eng exposure	.26	.05
			Spanish- like	Cumulative SPA Exposure	.05	.37
Sharp & Gathercole (2013)	45	55,32	Welsh-like nonwords (consonant scoring)	Current exposure	F(2,42)= 4.86	.013
			Welsh-like nonwords (cluster scoring)		F(2,42)= 3.38	.043
Summers et al. (2010)	62	[55.2 - 78]	Bilingual Spanish- and English-like nonwords	Age of exposure	-.15	> .05
					-.18	> .05
					-.26	> .05
					.16	> .05
					-.08	> .05

[5] 22 monolinguals and 22 bilinguals

[6] 50 monolinguals and 22 bilinguals

[7] 15 monolinguals and 42 bilinguals

[8] bilingual children with (N = 19) developmental language disorder

					-.40	< .05
Thordardottir (2017)	156	126,37	Icelandic- and nonword-like	Current exposure	.296	< .01
Thordardottir & Brandeker (2013)	84	60	Bilingual Icelandic- and English-like	Cumulative exposure	.141	.258
					.349	.005
Vender, Delfitto & Melloni (2018)	64[9]	125.81	Italian-like	monolingual vs bilingual NWR performance		

[9] 33 monolinguals and 31 bilinguals

Appendix C: Semi-structured interview subsections of parental concerns and language exposure features.

Biografia linguistica per i genitori

Intervistato: □ Madre □ Padre □Altro:_____ **Data di oggi:_____**

DATI DEL BAMBINO E DELLA SUA FAMIGLIA

Nome e cognome del

bambino:_____

Luogo di nascita: _____

Composizione della famiglia

Ruolo familiare rispetto al bambino (es mamma, papà…)		Età	Luogo di nascita (nazionalità)	Occupazione	Anni di scolarità (8= licenza media; 13= liceo/diploma; 18= laurea; >18= post lauream)
a					
b					
c					
d					
e					
f					

TAPPE EVOLUTIVE DEL BAMBINO

1. **A che età il bambino ha iniziato a camminare?** □< 15 mesi □> 16 mesi

2. **Quanti mesi aveva il bambino quando ha detto la sua prima parola in Italiano?**
   □< 15 mesi □16 – 24 mesi□> 25 mesi

3. **Quanti mesi aveva il bambino quando ha detto la sua prima parola nell'altra lingua**?
   □< 15 mesi □16 – 24 mesi□> 25 mesi

4. **A che età il bambino ha iniziato a mettere insieme le parole per formare frasi brev**i (es. “ancora latte”, “voglio acqua”, “scotta pappa”)?
   □< 24 mesi □25-30 mesi □> 30 mesi

5. **Si ricorda se suo figlio, utilizzava più gesti per farsi capire? Di che tipo? Fino a che età? _____**

INPUT LINGUISTICO

Mi può raccontare una settimana tipo del bambino?

Persone con cui trascorre il tempo quando non è a scuola		Tipo di lingua	Tempo(ore) durante la settimana	Tempo durante il week end
A				
B				
C				
D				
E				

_____
_____
_____

6. Prima di essere inserito a scuola, com'era la settimana tipo del bambino?

Persone con cui trascorre il tempo quando non è a scuola		Tipo di lingua	Tempo(ore) durante la settimana	Tempo durante il week end
A)				
B)				
C)				
D)				

Le capita di osservare se il bambino utilizza parole di due lingue diverse all'interno della stessa frase? □ Mai □ A volte □ Spesso

Nel complesso, siete soddisfatti di come il bambino parla l'Italiano**?**
□ Per niente soddisfatto □ Forse non soddisfatto □ Soddisfatto □ Completamente soddisfatto

Nel complesso, siete soddisfatti di come il bambino parla il_____(Lo)?
□ Per niente soddisfatto □ Forse non soddisfatto □ Soddisfatto □ Completamente soddisfatto

E' facile per la sua famiglia o i suoi amici avere una conversazione con suo figlio in Italiano?
□ Molto difficile □ A volte non è facile □ Abbastanza facile □ Molto facile

E’ facile per la sua famiglia o i suoi amici avere una conversazione con suo figlio in _____(Lo)?

□ Molto difficile □ A volte non è facile □ Abbastanza facile □ Molto facile

Pensa che il bambino stia perdendo il_____(Lo) in favore dell’italiano?

_____

_____

English translation:

Interviewed: □ Mother □ Father □ Other

CHILD AND FAMILY INFORMATION

Child name and surname:_____

Place of birth: _____

Family members:

Family role (e.g.: mum, dad, …)		Age	Place of birth (nationality)	Profession	Education (8= middle school diploma; 13= high school diploma; 18= master degree; >18= post lauream)
a					
b					
c					
d					
e					
f					

CHILD'S DEVELOPMENTAL PATHS

6. **When did your child start to walk?** □< 15 months □> 16 months

7. **How old was your child when s/he pronounced her/his first Italian word?**
   □< 15 months □16 – 24 months □> 25 months

8. **How old was your child when s/he pronounced his/her first word in her/his native language?**
   □< 15 months □16 – 24 months □> 25 months

9. **How old was your child when s/he started to combine two words?** (es. "Mum enough", "more water", "mum more")?
   □< 24 months □25-30 months □> 30 months

10. **Do you remember if your child used gestures to communicate? If yes, what kind of gestures? Until which age?** _____

LANGUAGE INPUT

Could you please describe the week-like activities of your child?

People spending time with child when not a school	Language(s) used	Weekdays time (hours)	Week-end time (hours)

A				
B				
C				
D				
E				

_____
_____
_____

Could you please describe the week-like activities of your child, before starting school?

People spending time with child		Language(s) used	Weekdays time (hours)	Week-end time (hours)
A)				
B)				
C)				
D)				

Does your child use words of two different languages in the same sentence?
□ Never □ Sometimes □ Often

Are you satisfied with your child's overall Italian proficiency?
□ Not at all satisfied □ Maybe satisfied □ Satisfied □ Completely satisfied

Are you satisfied with your child's overall proficiency in his/her native language?
□ Not at all satisfied □ Maybe satisfied □ Satisfied □ Completely satisfied

Is it easy for your family and friends to have a conversation in Italian with your child?
□ Very difficult □ Sometimes not easy □ Easy enough □ Very easy

Is it easy for your family and friends to have a conversation with your child in his/her native language?
□ Very difficult □ Sometimes not easy □ Easy enough □ Very easy

Do you think your child is losing his/her language proficiency in her/his native language in favour of Italian?

_____

_____

Appendix D: Current work's correlation matrix with age partialed out. Significance is set at alpha = .05 uncorrected.

	2	3	4	5	6	7	8	9
1 MB-CDI	.322	.305	.106	-.091	.224	-.057	.096	**0.472***
2 NWR words accuracy		**0.944****	**0.634****	.122	.103	-.226	-.096	**.713****
3 NWR phonemes accuracy			**.647****	.11	.062	-.152	-.212	**.766****
4 PPVT				.359	**.525***	**-.560***	-.092	**.634****
5 Maternal education					.279	-.227	-.394	-.114
6 Cumulative Italian exposure (%)						**-.851****	.322	.071
7 Age of Italian onset (months)							-.105	-.181
8 Current exposure to Italian (hours)								-.118
9 Parental concerns to Italian proficiency								1

*. Correlation is significant at the 0.05 level (2-tailed); **. Correlation is significant at the 0.01 level 2-tailed). NWR phonemes accuracy and parental concerns, and Age of Onset and Cumulative exposure correlations are significant when applying Bonferroni correction.

Appendix E: Current work's parametric correlation matrix. Significance is set at alpha = .05 uncorrected.

	2	3	4	5	6	7	8	9
1 MB-CDI	-.147	.351	.327	.18	-.09	.266	-.068	**.490***
2 Chronological age	1	-.384	-.226	**-,785****	0.142	**-.501***	.080	-.276
3 NWR (Whole-word scoring)		1	**.932****	**.663****	.059	.174	-.239	**.739****
4 NWR (phoneme scoring)			1	**,557***	.043	.038	-.165	**,780****
5 PPVT				1	.167	**.567***	-.408	**,594****
6 Maternal Education					1	.162	-.276	-.11
7 Cumulative Italian exposure (%)						1	**-.774****	.197
8 Age of Italian onset (months)							1	-.148
9 Parental concerns of Italian proficiency								1

*. Correlation is significant at the 0.05 level (2-tailed); **. Correlation is significant at the 0.01 level 2-tailed). NWR phonemes accuracy and parental concerns, NWR whole-word accuracy and parental concerns, Age and PPVT, as well as Age of Onset and Cumulative exposure correlations are significant when applying Bonferroni correction.

Appendix F: Re-analyses considering word length.

Whole-word scoring was estimated based on the 8 items for each word length. Phoneme scoring was based on the 35 phonemes in 2-syllable nonwords, 51 phonemes in 3-syllable items and 67 phonemes in 4-syllable nonword.

Looking at the effect of word length on NWR performance, we underline higher significant performance on 2-syllable stimuli than on both 3- (U=-4.933, p<.001) and 4-syllable (U=-3.584, p<.001) nonwords, and non-significant difference between 3- and 4-syllabic NWR performance (U=-1.389, p=.165), using whole-word scoring. With the phoneme scoring, we find similar trends: Although the difference between 2- and 3-syllable NWR performance was not significant (U=-1.041, p=.298), both 2- (U=-3.355, p=.001) and 3-syllabic stimuli (U=-2.616, p=.009) led to higher scores when compared to 4-syllable NWR performance. Thus, our results support the assumption that NWR performance decreases as nonwords' syllable length increases (Gibson et al., 2015; Summers et al., 2010).

Turning to our main results, the association between NWR performance and PPVT score was significant for 3- and 4-syllabic items, and weaker for 2-syllable items (particularly when phoneme scoring was used, where the estimate was close to zero). Although a higher-powered test would be necessary to be certain, these results could suggest that children enroll lexical knowledge to process longer nonwords, while performance for shorter stimuli seems less enhanced by lexical cues. We encourage further studies to focus on the role of lexical cues (e.g., word length and vocabulary scores) on NWR performance in bilingual children.

Parental concern was significantly correlated with NWR performance across word length and scoring definition, except for 2-syllabic stimuli calculated with phoneme scoring. Similarly, our conclusions regarding NWR and maternal education are not challenged either, with small to moderate non-significant estimates for all lengths and scoring modes.

Regarding the effect of nonwords' word length on NWR performance across MB-CDI scores in toddlerhood, word length does not impact the association between NWR and MB-CDI score when analyzing 2-, 3-, and 4-syllable stimuli separately, with small positive estimates for both whole-word and phoneme-based scoring. We underline children with higher (not significant but approaching significance) score on the MB-CDI outperformed those with scores lower than the 10th centile on 2-syllable nonwords using phoneme scoring. Those results enhance the assumption that children who obtained low MB-CDI scores in early childhood predict later phonemes repetition difficulties. Further studies should analyze such predictive value with larger sample size, in order to obtain results on a more representative bilingual population and increase statistical power. If our results are confirmed, a possible interpretation of this predictive value supports the idea that low expressive vocabulary size before the age of 3 is related with later phonemes' repetition difficulties, which might be due to phonological short-term memory difficulties. Further evidence will be welcome also on the potential role of phoneme scoring to detect significant phonemes' repetition difficulties compared to the whole-word scoring. Nevertheless, looking at our results, this interpretation is confirmed only by looking at different NWR performance across MB-CDI scores above or below the 10th percentile, because we report nonsignificant correlations between 2-syllable NWR performance and overall MB-CDI score of our whole sample. Further studies should investigate the predictive value of MB-CDI scores on NWR performance, looking at the effect of word length, as well as other sub-lexical cues, and different possible predictive values across children with MB-CDI scores' below or above the 10th centile.

[Table A near here]