The association between language exposure and non-word repetition performance in bilingual children: A meta-analysis

Gianmatteo Farabolini[a], Analí Taboh[bcd], Maria Gabriella Ceravolo[a], Federico Guerra[a]

[a]*Department of Clinical and Experimental Medicine, Università Politecnica delle Marche, Ancona, Italy*

[b]*Laboratorio de Neurociencia, Universidad Torcuato Di Tella, Buenos Aires, Argentina*

[c]*Departamento de Física, FCEyN, UBA and IFIBA, Conicet, Buenos Aires, Argentina*

[d]*Facultad de Filosofía y Letras, Universidad de Buenos Aires, Buenos Aires, Argentina*

Journals: Bilingualism: Language and cognition x 3, International Journal of Bilingual Education and Bilingualism x 5, International Journal of Bilingualism x 2, journal of child language

Abstract

Purpose: The purpose of this study was to explore the literature on the association between non-word repetition (NWR) performance and language exposure measures in bilingual children through a systematic review and meta-analysis. We studied the following research questions: is prior exposure to languages associated with non-word repetition performance in bilingual children populations? Is this association influenced by any variables?

Method: We entered selected search terms into two electronic databases: ERIC and Google Scholar. We conducted a blind literature review and a separate, random-effects model meta-analysis. We also tested potential moderators of the relationship between predictors and outcomes using sub-group meta-analysis. Last, we conducted publication-bias and sensitivity analyses.

Results: We identified 41 samples, comprising XXXX children,in a systematic review. We found significant association between NWR and language exposure using either cumulative or current exposure, non-language-like or language-like nonwords, whole-word or phoneme NWR scoring, and in children with typical language development. Nonsignificant associations were found using age of first exposure and NWR in bilingual children with atypical language development, even though results should be interpreted with caution given the small size of the studies included in this sub-group meta-analysis.

Conclusions: To our knowledge this is the first synthesis of the literature on the association between language exposure and non-word repetition performance in bilingual children. Our findings shed light on the contributions of different variables to NWR and highlight evidence to optimally design NWR tasks to assess bilingual language development.

Rationale

Bilingual children show heterogeneity in acquiring language, even greater than monolinguals. The main reasons for this seem to be linked to the complex mechanisms involved in acquiring two languages, as well as the amount of exposure to each of the languages the child is exposed to, from the environment (Carroll, 2017; Gatt & O'Toole, 2016). Bilingual children are exposed to (at least) one native language spoken in the family by one or both parents and a major language of a geographical area, which is the official language spoken outside the child's home (Farabolini et al., 2021). Language exposure (LE), is the prior experience to the languages a child is exposed to, and it shows an impact on language acquisition in children with both typical and atypical language development (Carroll, 2017).

Non-word repetition tasks

One research question is the disentanglement of variation in language acquisition patterns from atypical language development pathways. […] One of the most used clinical markers is non-word repetition (NWR), a neuropsychological task where children listen to a nonsense word that sounds like a real word in a specific language but has no meaning, and then they have to repeat it. A meta-analysis has shown this task identifies monolingual children with atypical language development with high accuracy (Estes et al., 2007).

Different research designs have been employed to develop nonwords. Language-like nonwords are those developed following the phonological constraints of a specific language; they are often created by changing some phonemes of real words (Engel de Abreu, 2011). non-language-like stimuli are developed respecting phonological rules of existing language(s), but they are often less strictly word-like: for example, authors often modulate nonwords' sub-lexical cues (e.g.: word length, phonotactic probability and wordlikness) to obtain this aim (Chiat & Polišenská, 2016; Thordardottir, 2017). Among sub-lexical cues, word length is the non-words syllabic range of a NWR list and it seems to impact NWR performance (Munson, Kurtz & Windsor, 2005; Brandeker & Thordardottir, 2015). To make some examples, /ˈnɑskət/ is an English-like nonword while /ˈsiˈpulɑ/ is non-language-like (Chiat & Polišenská, 2016), and /vopekɛt/ is a Ducth-like nonword and /lʊmikɑ/ is non-language-like (Boerma et al., 2015) and /kata'sepo/ is an Italian-like nonword (Dispaldro, Leonard & Deevy, 2013). Crosslinguistic nonwords are a subtype of non-language-like that which are developed respecting the phonological constraints of a set of languages, with the aim of maximally reducing the effect of experience and proficiency in different languages spoken by bilingual children on NWR tasks (Armon-Lotem et al., 2015).

Two notation systems have been used in the literature to calculate NWR performance. […].

Non-word repetition in bilingual children

Looking at bilingual development, working memory and verbal short-term memory tasks are often used in bilingual language assessment and mixed evidence has been found on its relationship with language scores (Monnier et al., 2021). Among such tasks, NWR is studied for two main purposes. […].

Different authors raised debates on the impact of lingual status (monolingualism vs bilingualism) on NWR performance and related variables that might play a role on such relationship (Armon-Lotem & Meir, 2016; Gutiérrez-Clellen & Simon-Cereijido, 2010).For example, the nature of NWR stimuli (e.g.: crosslinguistic vs language-specific) might have an influence on NWR performance in bilinguals (Chiat & Polišenská, 2016), as well as non-words' sub-lexical cues (e.g.: word length, that is the syllabic range of NWR list ,or wordlikness, that is how nonwords sounds like a real word of a certain language; Munson, Kurtz & Windsor, 2005; Brandeker & Thordardottir, 2015). Then, mixed evidence has been found on the association between language exposure to a specific target language and NWR performance (Bonifacci et al., 2018; Core et al., 2017; Thordardottir, 2017).

Non-word repetition performance and language exposure

Evidence supports the claim that prior exposure to a specific language partially explains heterogeneity in receptive and expressive scores, as well as in non-word repetition performance (REF REF), but mixed evidence has been found on the relationship between NWR and LE. Different hypotheses support either the independence or the relationship between NWR and LE.

On one hand, […] this task might be helpful to identify atypical language development in bilingual children, with prior language exposure which should at most minimally affect NWR performance. […] Indeed, NWR could be helpful to disentangle low langauge assessment scores due to lower exposure in the language the assessment is carried out, from those due to language difficulties.

On the other hand, […] the task requires to storage, retrieve and reproduce a meaningless sequence of phonemes. Even though NWR is not a linguistic task that mainly assesses a language domain, prior language exposure should play at least a low impact on NWR. The impact of prior language exposure to NWR might be mediated by non-words' sub-lexical cues […]. From this perspective, different authors aimed to develop non-language-like non-words in order to develop non-words minimally influenced by prior language exposure (Chiat & Polišenská, Thordardottir, 2017). Mixed evidence has been found on the relationship between non-language-like NWR performance and language exposure (Tuller et al., 2019; Armon-Lotem, 2017).

Different language exposure measures are frequently used to study its association with NWR, namely age of first exposure, current exposure, and cumulative exposure. […].

Potential moderators can affect the relationship between NWR & LE.

First, language exposure might be differently related in children with and without atypical language development. […].

Second, chronological age might play a role in the association between NWR & LE. […].

To study the optimal research design for NWR research designs to assess bilingual language development, it might be relevant to understand the impact of language exposure on NWR performance, and the weight of the related variables on such association.

The current study

We carried out a systematic review and meta-analysis addressing the following research questions:

1) Is prior exposure to language associated with non-word repetition performance in bilingual children populations?
2) Which are the variables related to bilingual language development that affect the association between NWR and language experience?

   Is the association between NWR and language exposure influenced by:

   a) the measure of language exposure (cumulative exposure, current exposure, or age of first exposure)?
   b) the type of stimuli in the NWR task (non-language-like vs language-specific)?
   c) the system used to score the NWR task (whole-word vs phoneme scoring)?

d) language development (typical or atypical)
e) participants' chronological age (in toddlers, pre-scholars and scholars)?

Methods

Our research design follows PRISMA guidelines for systematic reviews and meta-analyses (Liberati et al., 2009). The current study was registered in the "International Prospective Register of Systematic Reviews" (PROSPERO) in 2020 (CRD42020173573). The first research question of the current systematic review studies the association between two variables, while the second analyses whether the meta-analysis main effect is influenced by any moderators.

Systematic review

Search protocol

We used the open-access databases Google Scholar and ERIC for database searching. Independent search was carried out by the first and the second authors. For the Google Scholar database we employed the following search keywords: […]. Study selection was carried out following these steps: abstract retrieval, abstract screening, full text retrieval and full text screening. Once the literature search and screening were finished, we compared our inclusion decisions and reached an agreement after a consensus process; when consensus was not reached by the first and the second author, we asked for the last author advice to take a decision. For papers initially included by one author but not by the other, full text screening and data extraction were carried out by both. We also added research articles from personal readings, as well as data received from a research team working on nonword repetition performance in bilingual children (Pérez-Navarro et al., 2020): from their work, we include one unpublished article in our systematic review.

From included papers, we extracted data about participants, language exposure measures, NWR research design, NWR scoring system, and statistical analysis. […].

Inclusion and exclusion criteria

To be included in the current systematic review, experimental data had to be related to bilingual children, i.e. children [0-18 years] who are exposed to at least two languages during their lifespan. We included studies that employed NWR task (but not word learning or sentence imitation assessment tools) and reported statistical results for the association between NWR and language exposure measures. We excluded works reporting results about the association between NWR task developed following constraints of a language (e.g.: Spanish) and language exposure measured looking at a different language (e.g.: English). Multi-comparison results not reporting the single effect of language exposure measure on NWR scores were excluded as well. In addition, in order to follow an inclusive approach, we decided to include data from the comparison between monolingual and bilingual children on a NWR task since these results report the difference between groups differing for prior language exposure. .

Data extraction

After consensus' processes on included studies, the first two authors carried out a blind data extraction, and then they converge and discuss their data extraction. For included articles, we reported the sample size, the language development of the children (typical vs atypical) and their lingual status (only bilingual children or mixed populations), nonwords type (cross-linguistic, language-like, non-language-like or mixed), the scoring systems employed (phoneme and/or whole-word scoring) and language exposure (e.g.: cumulative exposure, current exposure, age of onset, length of exposure). We decided to not report correlational results based on the phonological constraints of a certain language to develop NWR and exposure measure on a different language (e.g.: results on English-like NWR and Spanish exposure Gathercole & Masoura, 2005; Pérez-Navarro, Molinaro, & Lallier, 2020; Parra, Hoff & Core, 2011; Summers et al., 2010). […].

Risk of bias

We developed a list of study risk-of-bias variables after consulting literature and considering methodological issues which can affect the quality of the information derived from a study, regarding the research question of the current work. The following study characteristics were assessed: (a) representativeness of the exposed cohort; (b) published vs unpublished data; amount of nonwords administered; (c) bilingual status (if participants share only the major language or also the native one); d) data bearing on native or major language (language exposure and non-words based and developed following the native rather than the major language); e) amount of nonwords administered, f) published or unpublished articles (see Supplemental Material S2 for further information). We did not exclude studies based on risk of bias. Rather, we took an inclusive approach to study selection, in order to maximize the literature of our systematic review and meta-analysis and still collect our data related to risk of bias.

Data processing

[…].

Meta-analysis

We conducted a random-effects model meta-analysis for each individual predictor using Review Manager 5.4 (The Cochrane Collaboration, The Nordic Cochrane Centre, Copenhagen, Denmark). We acknowledged significance at $p < .05$. .

Main effects

We reported the main effects of each meta-analysis including standard errors in our software, as well as our Fisher's z as Odd ratios. We used such analysis for both association and comparison (monolingual vs bilingual) research designs. We used the same methodology for all the sub-group meta-analysis carried out. We found that some articles also reported the association between NWR developed in a certain language and language exposure in a different language (e.g.: French-like NWR and English language exposure) (Brandeker & Thordardottir, 2015; Gathercole & Masoura, 2005; Li'el, 2017; Parra et al., 2011; Pérez-Navarro et al., 2020; Summers et al., 2010): we decided to exclude such results given that they are not related to our research questions. However, we also reported in the results section the main effect including such data.

Heterogeneity

The statistical significance of heterogeneity was analysed using a chi-square test of the Q value and estimated the magnitude of heterogeneity using the I^2 value (Borenstein et al., 2009). The interpretation of I^2 was low (< .25), moderate (.50), or high (.75) (Higgins et al., 2003). We report heterogeneity of each sub-group meta-analysis on each related forest plot.

Moderation analysis

We tested potential moderators of the relationship between predictors and outcomes using sub-group meta-analysis. Recommendations indicate that moderation analyses are appropriate when the following conditions are met: presence of moderate–high heterogeneity ($I^2 > .25$) and a minimum of eight studies for sub-group meta-analysis (Borenstein et al., 2009). […].

Publication bias and sensitivity analysis

We carried out publication-bias and sensitivity analyses to evaluate the validity and robustness of meta-analysis findings. We assessed publication bias in meta-analyses by examining funnel plots for asymmetry and applying the Newcastle-Ottawa Scale (Stang, 2010), as well as conducting sub-group analysis (see Supplemental materials S2 for further information). We assessed sensitivity by exploring the effects of removing each individual study on our meta-analysis and sub-group analysis results (Fisher et al., 2016). We only report sensitivity results which change the main effect.

Statistical analysis

We choose a random effect model to carry out all our quantitative integrations. We converted the results extracted from the articles into Fisher's z scores and we calculated standard errors for each study. Unfortunately, using the open-access Review Manager softaware, it is impossible to calculate a meta-analysis reporting correlation results as effect sizes (Fisher, 2017). We decided to insert Fisher's z scores as our odd ratio and its related standard error.

Results

Quality assessment

See Supplemental Material S1 for the PRISMA flow-diagram detailing search results and numbers of records that we excluded for various reasons. We found a total of 41 articles reporting data on our research question (see Table 1). Out of those 41, 30 have been published in peer-reviewed journals as experimental studies, and thus have been evaluated as such. Among the remaining eleven works, there are one study under submission

(Pérez-Navarro et al., 2020), a commentary on a special issue (Armon-Lotem, 2017), a study presented in conference proceedings (Core et al., 2017), four PhD dissertation (Eliseeva et al., 2018; Kalak, 2020; Oberg, 2020; Orellana, 2020), and master theses (Huls, 2017; Li'el, 2017; Limacher, 2019).

[Table 1 near here]

Studies' sample sizes are heterogeneous, with a median of 60 children (mean= 62.4, range=16-266, total= 2244). Participants were recruted from toddlerhood to high school (mean age=75.68 [22.78-160.8], SD=32.44; median age=71.45). XX studies reported data on samples of bilingual children only, while the remaining XX also reported data on monolingual peers in order to study different NWR performance across lingual status (monolingual vs bilinguals[…]).XX studies reported results on both bilinguals only and monolingual vs bilingual status(Antonijevic et al., 2019; Haman et al., 2017; Limacher, 2019).. The majority of studies included children with typical language development, others report both children with typical and atypical language development (Antonijevic et al., 2019; Tuller et al., 2018), while others reported results for children with and without developmental language disorders separately (Boerma et al., 2015; de Almeida et al., 2017; Li'el, 2017; Talli, & Stavrakaki, 2020).

Some studies presented samples composed by only bilingual children being exposed to more than one language during childhood, others included also monolingual children in their studies and reported monolingual-bilingual comparison results . Twenty-two studies reported data on samples composed of children exposed to the same set of languages ([…]).

Different languages were spoken by bilingual participants. The majority of studies reported English as major language and Spanish […], Icelandic (Thordardottir, 2017), Korean, Chinese or other South Asian languages […], French […], Polish (Haman et al., 2017), Welsh (Sharp & Gathercole, 2013), Greek (Gathercole & Masoura, 2005) or Italian (Eliseeva et al., 2018) as native language. Other publications included children having French […], English […], Italian […], Icelandic (Thordardottir & Juliusdottir, 2013), Australian English (Li'el, 2017) as their major language and a range of different native languages. Two studies included Russian-Hebrew bilingual children […], one study included Turkish-Dutch bilinguals (Schraeyen et al., 2018), one Basque-Spanish bilinguals (Pérez-Navarro et al., 2020) and one research article collected data on a sample composed by Chinese-French bilingual children (Pierce et al., 2015). Two studies enrolled bilingual children in Luxembourg, having either Portuguese or Brazilian Portuguese (Engel de Abreu et al., 2013) or different languages (Engel de Abreu, 2011) as their native language and Luxembourgish as their major language. It should be noted that Luxembourg is officially trilingual: French and German are generally learned at school (Engel de Abreu, 2011).

Different language exposure measures have been used by researchers: cumulative exposure ([…]), current exposure ([…]), age of onset ([…]). We In addition, 19 studies analysed the association between language exposure and NWR accuracy comparing performance between monolingual and bilingual peers ([…]).

Nonword stimuli included in the current systematic review were developed following either a language-like or a non-language-like research design. Some studies administered two different lists of language-like nonwords specific to each of the languages spoken by the children (e.g.: both Spanish-like and English-like nonwords for Spanish-English bilinguals) ([…]), while others reported results from both language-like and non-language like nonwords ([…]); for these studies, some authors reported single results from each NWR type ([…]), while others reported combined NWR performance using nonwords developed with different theoretical frameworks ([…]). Nonword stimuli also differ in word length, ranging from 1- to 7-syllable items. NWR accuracy was calculated using whole-word scoring ([…]), phoneme scoring ([…]) or both of them ([…]). Additionally, some studies used language-like stimuli constructed according to the phonological constraints of the standard variety of a language, while children were exposed to a geographical variety of that language: for example, North America English-like nonwords are used for bilingual children having Australian English as their major language (Li'el, 2017).

Risk of bias

A checklist of desirable study characteristics is given in Supplemental Material S2. We found twenty-eight studies including more than 50 participants.Thirty-three studies are published in international peer-reviewed journals. Twenty studies reported data on bilingual participants only, while twenty-two reported results comparing NWR performance between monolinguals and bilinguals. Eight studies had used and developed NWR and LE in the participants' native languages, while twenty-seven studies used the major language. Thirteen studies used an amount of nonwords between 16 and 24, and fourteen studies used either less than 8 or more than 40. Finally, fourteen studies had a syllabic range of [1-5], [1-4] or [2-5], and fifteen used stimuli with a syllabic range between 3 and 2, excluding 6-syllable nonwords and above.

Quantitative integration

We carried out a Random effect meta-analysis on the selected studies. Across the forty-one articles, the main effect revealed a positive and significant association between NWR performance and language exposure (OR: 1.23 [1.18, 1.28]; $p < .00001$), with medium heterogeneity across studies' results (I^2=44%) . [...].

[Figure 1 near here]

[Figure 2 near here]

We then carried out sub-group meta-analysis to study the effect of different variables which can affect the main effect (see Table 2 for summary findings).
Regarding the impact of language exposure, while positive and significant associations were found between NWR and both cumulative ([...]) and current ([...]) exposure, age of onset seems to not be related with NWR performance ([...]).
Concerning NWR stimuli type, the association remains positive and significant using both non-language-like ([...]) and language-specific ([...]) nonwords. Looking at the effect of NWR scoring system, similar results were reported using whole-word ([...]) or phoneme ([...]) scoring.
Then, considering studies reporting data only on children with typical language development, the association results did not change ([...]). Looking at studies on children with atypical language development only, the association is not significant [...]; the result does not change when conducting sensitivity analysis. This result should be taken with caution since it does not respect recommendations (four studies included in the sub-group analysis).

[Table 2 near here]

We finally looked at the effect of selected bias on meta-analysis.
Main effect meta-analysis did not change considering studies with more or less 50 participants ([...]). The main effect does not change considering evidence of published

articles. Looking at unpublished articles, the sub-group analysis approach significance ([…]). Sensitivity analysis revealed the association is not anymore significant ([…]removing Elisseva 2018; […] removing Huls, 2017; […] removing Oberg, 2020). The main effect does not chage when considering studies including either bilingual participants only or comparing bilingual and monolingual performance ([…]), as well as across amounts of nonwords administered and its syllabic range (see Supplemental material XX for further information).

Discussion

Our meta-analysis investigated the association between prior language exposure to languages and NWR performance. and we found a significant correlation. We then carried out sub-group meta-analysis to further examine which variable might affect the association.

We found that both current and cumulative exposure are associated with NWR performance, while, interestingly, age of first exposure to a language seems to not affect NWR score. A possible interpretation of this finding is that the former measures are more representative of and more closely associated to language development in bilingual children than the latter. One possible interpretation is that […].

It has to be underlined that almost the totality of articles employed quantitative language exposure measures. Quality of input features (e.g.: *XX*) should be taken into account to analyse the role of language exposure and experience on language development (Farabolini et al., 2021; Hoff, 2020).

NWR stimuli type seems to not affect the association between language exposure and NWR. Indeed, similar results have been found for studies using either language-like or non-language-like stimuli. These results do not support the assumption that non-language-like stimuli should be preferred because they maximally reduce the impact of language exposure to a target language on NWR (Armon-Lotem et al., 2015; Chiat & Polišenská, 2016). This result does not discourage using non-language-like, but shed light on some considerations and both research and clinical implications.

First, our evidence underlines that it seems really challenging to develop language assessment tools which are not related to prior language exposure in bilingual children. However, we encourage to develop nonwords that are optimally the more independent from prior language exposure, considering […].

Concerning clinical implications, in line with previous literature (Armon-Lotem, de Jong & Meir, 2015), NWR should be used within other assessment tools. Moreover, language exposure features should be taken into account even when non-language-like nonwords are employed.

Looking at the effect of the NWR scoring system, we found it does not seem to affect the association between NWR and language exposure. This finding is in line with previous results (Dispaldro et al., 2013; Estes et al., 2007; Farabolini et al., 2021), supporting the hypothesis that, even though more detailed information might be collected with phoneme scoring on phonological and articulatory skills (Piazzalunga et al., 2019), both NWR scoring systems can be employed to study the impact of language exposure on NWR performance.

We found the association between NWR and language exposure is not significant in children with atypical language development, […].

Finally, looking at the risk of bias assessment, we found the main effect differ for unpublished articles. The weak significance of the association between NWR & LE might suggest there is a bias of publication of negative results. Such bias might be present in both research designs including bilingual participants only or monolingual too.

Limitations

Several limitations have to be reported regarding included articles in the current review.

First, we excluded both studies that were not available in English, Spanish, French or Italian, and studies for which an effect size could not be calculated.

Second, there are sample differences across studies that might arise heterogeneity in our work. For example, as mentioned above, we have included studies reporting data on only bilingual children and on both monolingual and bilingual children. Even though these two groups of studies use two different research designs, we believe they both address the question about the relationship between prior exposure to a target language and NWR accuracy. Our hypothesis is confirmed by sub-group meta-analysis revealing the main effect does not change considering studies including either bilinguals only or comparing monolingual vs bilingual scores. However, we suggest to take into account our evidence also include data from monolingual samples.

We also have to highlight methodological differences in research designs between included studies.

For example, in our risk of bias assessment, sub-group analysis revealed there are not any effects due to the amount of or the syllabic range of nonwords administered. However, we highlight these two factors shape nonwords' features and should be taken into account when using NWR. At least, we underline hetereogenity in both the amount of nonwords administered to participants ranged from 8 (Bonifacci et al., 2018) to 108 (Sharp & Gathercole, 2013) (M=30.48), as well as the range of nonword length in syllables varied from [1-2] to [3-5] in lists with shorter stimuli and from [1-7] to [2-6] in lists with longer stimuli. Mixed evidence has been found in the literature on the effect of nonword length on NWR performance. While some report NWR performance decreases as nonword syllable length increases (Chiat & Polišenská, 2016; Gibson et al., 2015; Summers et al., 2010), others report the absence of a significant effect (Farabolini et al., 2021). Further research should study the role of sub-lexical cues (e.g.: stimuli's word length) on NWR performance.

Then, there is one other limitation that should be considered when interpreting the results presented in this work, which is that it includes data of NWR tasks built on different language-specific phonological constraints. As a consequence, stimuli with similar characteristics might still differ greatly across languages. For example, disyllabic items in French can be "*vapagne"* or *"chansette"* (Pierce et al., 2015), in Italian "*bofo*" or "*simi*" (Dispaldro et al., 2013), and in English "nagi" or "lumi" (Chiat & Polišenská, 2016). Such differences (e.g.: number of phonemes required to reproduce bisyllabic stimuli, articulatory complexity) might affect NWR performance and the related association with LE.

In addition, similarity between the languages the child is exposed to might influence our research question. For example, we imagine that bilinguals exposed to langauges that are similar (e.g.: French and Spanish) , could benefit more of the prior exposure received in the native language and "use it" to acquire major language (e.g.: language transfer) rather than children exposed to languages with lower similarity (e.g.: Mandarin and English). Among others, articulatory complexity, language-specific phonological development and NWR research design are variables that should be considered in studies on NWR performance in bilingual children.

Last but not least, we underline all included studies were conducted in Western, educated, industrialized, rich, and democratic (WEIRD) countries. It should be interesting collecting data in non-WEIRD countries to highlight the weight of social and cultural differences in developmental pathways (Muthukrishna et al., 2020). Interestingly, a recent work revealed NWR is not associated to infant-direct speech from mothers and relatives to children, in a community of monolingual kids growing in Amazonian villages (hidden for blind review, 2020).

Conclusions

As has been conveyed throughout this work, heterogeneity is the keyword to describe differences both between and within bilingual populations, considering the languages spoken and the geographical areas, as well as each individual's linguistic background. Nonetheless, research and clinical communities are working to improve multilingual language assessment for children exposed to more than one language. This review and meta-analysis, which included studies on bilingual children with a wide range of languages spoken, geographical areas and chronological ages, shows that NWR performance is significantly associated with the prior language exposure received, specially measured by cumulative and current exposure. Further studies should focus on this association in bilingual children with atypical language development, since we found weak association in this population. Our findings encourage the use of NWR tasks on bilingual children, keeping in mind language exposure plays a core role in NWR accuracy on bilingual children. Given this task's potential for bilingual language assessment, we hope this work will contribute to better and deeper understand the cognitive and linguistic mechanisms involved in NWR tasks.

Disclosure of interest

The authors have no conflicts of interest to declare.

Data availability statement.

We present descriptive statistics with data collected on our sample.

Author contributions

References

Armon-Lotem, S., de Jong, J., & Meir, N. (Eds.). (2015). *Assessing multilingual children: Disentangling bilingualism from language impairment*. Multilingual matters. https://www.narcis.nl/publication/RecordID/oai:dare.uva.nl:publications%2F4d4794cb-3ec1-4413-b349-2d591da9a2a3.

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.

Bonifacci, P., Atti, E., Casamenti, M., Piani, B., Porrelli, M., & Mari, R. (2020). Which Measures Better Discriminate Language Minority Bilingual Children With and Without Developmental Language Disorder? A Study Testing a Combined Protocol of First and Second Language Assessment. *Journal of Speech, Language, and Hearing Research*, 1-18. https://doi.org/10.1044/2020_JSLHR-19-00100.

Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2009). Introduction to meta-analysis. Chichester, United Kingdom: Wiley.

Bortolini, U., Arfé, B., Degasperi, L. Deevy, P. & Leonard, L. B., (2006). Clinical markers for specific language impairment in Italian: the contribution of clitics and non‐word repetition. *International Journal of Language & Communication Disorders,* 41, n. 6 : 695–712.

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.

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), 1179-1189. https://doi.org/10.1044/2016_JSLHR-L-15-0293.

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.

Conti‐Ramsden, G., Botting, N., & Faragher, B., (2001). Psycholinguistic Markers for Specific Language Impairment (SLI). *Journal of Child Psychology and Psychiatry,* 42(6) 741–48.

dos Santos, C., & Ferré, S., (2016): A Nonword Repetition Task to Assess Bilingual Children's Phonology. *Language Acquisition*, DOI: https://doi.org/10.1080/10489223.2016.1243692

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), 177-195. https://doi.org/10.1044/1092-4388(2007/015).

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

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), 117-123. https://doi.org/10.1080/13670050.2016.1179926.

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.

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.

Lakens, D. (2013). Calculating and reporting effect sizes to facilitate cumulative science: a practical primer for t-tests and ANOVAs. *Frontiers in psychology*, *4*, 863. https://doi.org/10.3389/fpsyg.2013.00863.

Liberati, A., Altman, D. G., Tetzlaff, J., Mulrow, C., Gøtzsche, P. C., Ioannidis, J. P., ... & Moher, D. (2009). The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration. *Journal of clinical epidemiology*, *62*(10), e1-e34.

Piazzalunga, S., Previtali, L., Pozzoli, R., Scarponi, L., & Schindler, A. (2019). An articulatory-based disyllabic and trisyllabic Non-Word Repetition test: reliability and validity in Italian 3-to 7-year-old children. *Clinical Linguistics & Phonetics.* 33(5), 437-456. https://doi.org/10.1080/02699206.2018.1542542

Pérez-Navarro, J. J., Molinaro, N., & Lallier, M. (2020). The influence of amount of exposure on bilingual language development: a longitudinal study of Basque-Spanish preschoolers. https://doi.org/10.17605/OSF.IO/XG39C.

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*, 1–23. https://doi.org/10.1080/13670050.2017.1358695.

Tables and figures legends

Table 1: Participant and methodological characteristics of included records and unpublished reports. Notes: L1=native language, L2=dominant language in a specific geographical area.

Figure 1: Forest plot of the association between nonword repetition and language exposure.

Figure 2: Funnel plot of the association between nonword repetition and language exposure.

Table 2: Summary of evidence from sub-group meta-analysis.

Table 1

Figure 1

Study or Subgroup	log[Odds Ratio]	SE	Weight	Odds Ratio IV, Random, 95% CI
Altman et al., 2014	0.304	0.127	2.9%	1.36 [1.06, 1.74]
Antonijevic et al., 2019	-0.11	0.13	2.8%	0.90 [0.69, 1.16]
Armon-Lotem, 2017	0.37	0.132	2.8%	1.45 [1.12, 1.88]
Barbosa, Jiang & Nicoladis 2017	0.562	0.115	3.2%	1.75 [1.40, 2.20]
Bellocchi et al., 2017	0.07	0.11	3.3%	1.07 [0.86, 1.33]
Boerma et al., 2015	0.139	0.138	2.6%	1.15 [0.88, 1.51]
Bonifacci et al., 2018	0.085	0.089	3.9%	1.09 [0.91, 1.30]
Brandeker & Thordardottir, 2015	0.136	0.149	2.4%	1.15 [0.86, 1.53]
Buac, Gross & Kaushanskaya, 2016	0.334	0.112	3.3%	1.40 [1.12, 1.74]
Chiat & Polišenská, 2016	0.129	0.16	2.2%	1.14 [0.83, 1.56]
Core, Chaturvedi & Martinez-Nadramia, 2017	0.06	0.099	3.6%	1.06 [0.87, 1.29]
de Almeida et al., 2017	0.14	0.184	1.8%	1.15 [0.80, 1.65]
Duncan & Paradis, 2016	0.04	0.118	3.1%	1.04 [0.83, 1.31]
Eliseeva, Ziffer & Marini, 2018	0.352	0.13	2.8%	1.42 [1.10, 1.83]
Engel de Abreu, 2011	0.345	0.156	2.3%	1.41 [1.04, 1.92]
Engel de Abreu et al., 2013	0.542	0.164	2.1%	1.72 [1.25, 2.37]
Farabolini et al., 2020	0.016	0.25	1.2%	1.02 [0.62, 1.66]
Gibson et al., 2015	0.313	0.143	2.5%	1.37 [1.03, 1.81]
Haman et al., 2017	0.545	0.08	4.2%	1.72 [1.47, 2.02]
Huls, 2017	0.218	0.151	2.4%	1.24 [0.93, 1.67]
Kohnert et al., 2006	0.378	0.12	3.0%	1.46 [1.15, 1.85]
Lee & Gorman, 2013	0.352	0.136	2.7%	1.42 [1.09, 1.86]
Li'el, 2017	0.002	0.25	1.2%	1.00 [0.61, 1.64]
Limacher, 2019	0.106	0.267	1.0%	1.11 [0.66, 1.88]
Masoura & Gathercole, 2007	0.218	0.114	3.2%	1.24 [0.99, 1.55]
Namazi & Thordardottir, 2010	0.318	0.192	1.7%	1.37 [0.94, 2.00]
Parra, Hoff & Core 2011	0.065	0.162	2.2%	1.07 [0.78, 1.47]
Pérez-Navarro, Molinaro, & Lallier (2020)	0.023	0.125	2.9%	1.02 [0.80, 1.31]
Pierce et al., 2015	0.293	0.158	2.2%	1.34 [0.98, 1.83]
Schraeyen et al., 2018	0.472	0.169	2.1%	1.60 [1.15, 2.23]
Sharp & Gathercole, 2013	0.263	0.154	2.3%	1.30 [0.96, 1.76]
Summers et al., 2010	0.176	0.127	2.9%	1.19 [0.93, 1.53]
Thordardottir, 2017	0.069	0.062	4.8%	1.07 [0.95, 1.21]
Thordardottir & Brandeker, 2013	0.177	0.127	2.9%	1.19 [0.93, 1.53]
Thordardottir & Juliusdottir, 2013	0.654	0.277	1.0%	1.92 [1.12, 3.31]
Tuller et al., 2019	0.104	0.023	5.8%	1.11 [1.06, 1.16]
Vender, Delfitto & Melloni, 2018	0.178	0.128	2.9%	1.19 [0.93, 1.54]
Total (95% CI)			**100.0%**	**1.25 [1.18, 1.32]**

Heterogeneity: Tau² = 0.02; Chi² = 83.97, df = 36 (P < 0.0001); I² = 57%
Test for overall effect: Z = 7.39 (P < 0.00001)

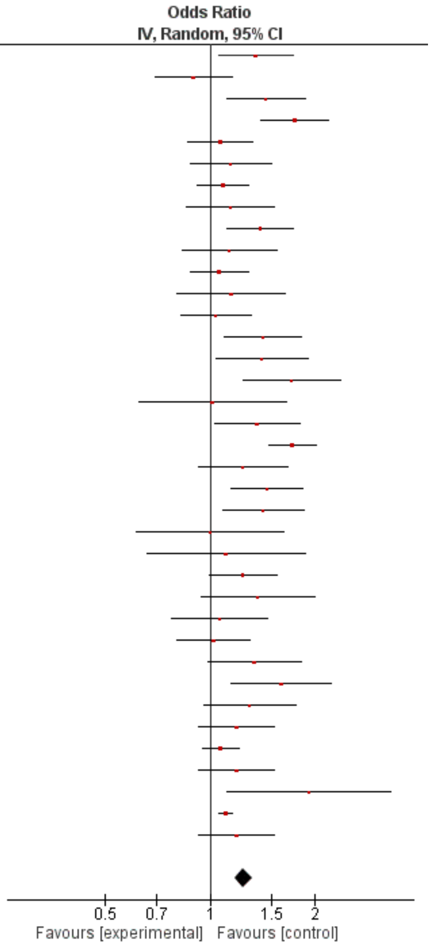

Figure 2

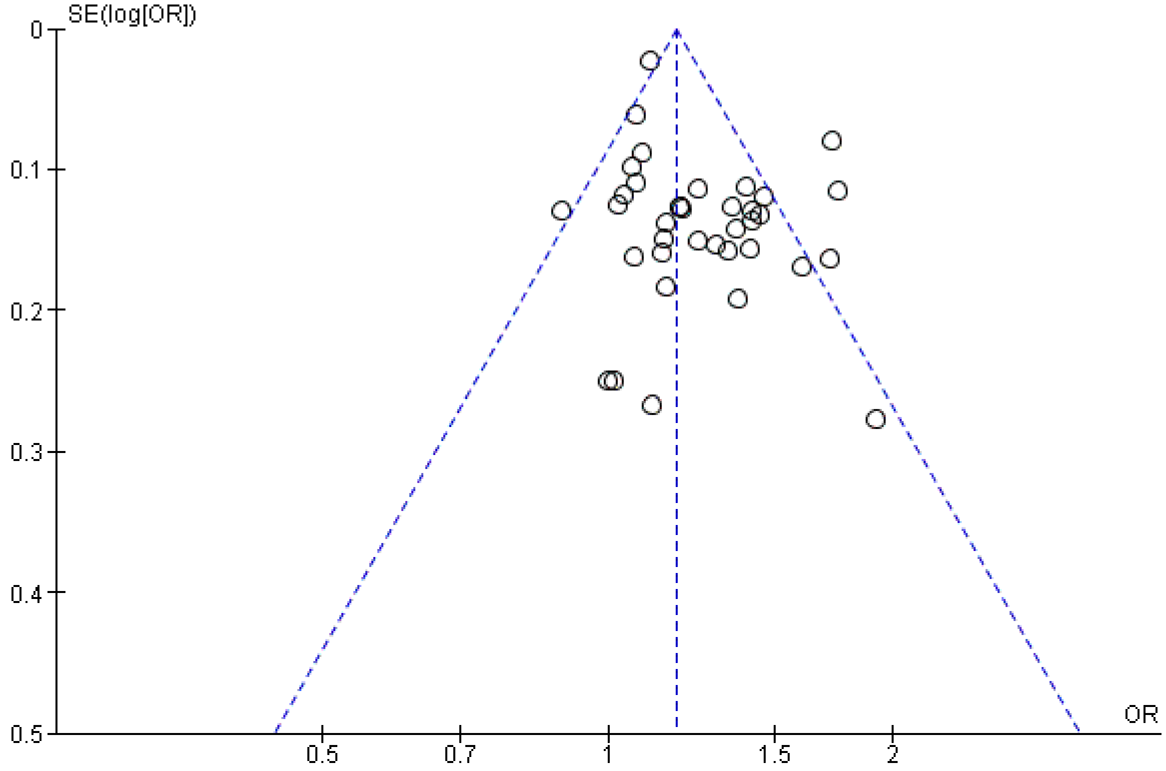

Table 2

Variables	N studies	N of participants	Results (OR)	Significance
Nonlanguage-like NWR & LE	7	714	1.17 [1.02 - 1.34]	.02
Language-like NWR & LE	30	2037	1.22 [1.15, 1.31]	<.00001
NWR & cumulative exposure	9		1.17 [1.09, 1.25]	<.0001
NWR & current exposure	6		1.20 [1.06, 1.35]	.004
NWR & age of first exposure	8		0.97 [0.79, 1.21]	.82
NWR & LE using whole-word scoring	21		1.25 [1.14, 1.37]	<.00001
NWR & LE using phoneme-scoring	16		1.19 [1.11, 1.26]	<.00001
NWR & LE in sample with typical lang. dev.	35		1.27 [1.20, 1.35]	<.00001
NWR & LE in sample with atypical lang. dev.	3		0.98 [0.80, 1.21]	.88
NWR & LE in published studies	32		1.26 [1.18, 1.34]	<.00001
NWR & LE in unpublished studies	5		1.18 [1.03, 1.36]	.02