Bilingualism: Language and Cognition

cambridge.org/bil

Review Article

Cite this article: Farabolini G, Taboh AR, Ceravolo MG, Guerra F (2023). The association between language exposure and nonword repetition performance in bilingual children: A systematic review and meta-analysis. *Bilingualism: Language and Cognition* **26**, 621–635. https://doi.org/10.1017/S1366728922000906

Received: 15 December 2021
Revised: 11 November 2022
Accepted: 20 November 2022
First published online: 28 December 2022

Keywords:
nonword repetition; language exposure; bilingualism; systematic review; meta-analysis; children

Address for correspondence:
Gianmatteo Farabolini, Department of Experimental and Clinical Medicine, Università Politecnica delle Marche, Ancona, Italy.
E-mail: g.farabolini@pm.univpm.it; gianmatteo.farabolini@gmail.com.

The association between language exposure and nonword repetition performance in bilingual children: A systematic review and meta-analysis

Gianmatteo Farabolini[a], Analí R. Taboh[b,c,d], Maria Gabriella Ceravolo[a] and Federico Guerra[a]

[a]Department of Experimental and Clinical 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 and [d]Facultad de Filosofía y Letras, Universidad de Buenos Aires, Buenos Aires, Argentina

Abstract

We conducted a systematic review and meta-analysis on the association between nonword repetition (NWR) and language exposure in bilingual children and explored whether the association is influenced by other variables. We performed a blind literature review on ERIC and Google Scholar, a random-effects model meta-analysis and subgroup analyses to test potential moderators. Out of 822 screened articles, we identified 24 works including 1399 children. Significant associations were found using either cumulative or current exposure, language-like nonwords, phoneme NWR scoring, in children with typical language development. Nonsignificant associations were found in studies either using age of first exposure, on children older than six, with less than 50 participants, using NWR lists containing 16-24 nonwords or with participants having different native languages. Weak associations were found when considering whole-word scoring or gray literature. We highlight the contributions of different variables to NWR, and evidence to optimally design NWR for bilingual language assessment.

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 (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, Caselli, Rinaldi, & Cristia, 2021). Language exposure is the prior exposure to each of the languages a child is exposed to, and it has an impact on language acquisition in children with both typical and atypical language development (Carroll, 2017).

Nonword repetition tasks

Given the heterogeneity in bilingual language acquisition, there is a growing body of research on how to disentangle variation in acquisition patterns from atypical language development pathways. To address this question, different authors have worked on clinical markers, where evidence has shown significantly lower performance in monolingual children with atypical language development if compared to children with typical language development (Bortolini, Arfé, Caselli, Degasperi, Deevy, & Leonard, 2006; Conti-Ramsden, Botting, & Faragher, 2001). One of the most used clinical markers is nonword repetition (NWR), a neuropsychological task where children listen to a nonsense word that sounds like a real word but has no meaning, and then they have to repeat it. Recent meta-analyses have shown this task identifies monolingual (Graf Estes, Evans, & Else-Quest, 2007) and bilingual (Ortiz, 2021; Schwob, Eddé, Jacquin, Leboulanger, Picard, Ramos Oliveira, & Skoruppa, 2021) children with atypical language development.

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 the phonological rules of one or two existing languages, but they are often less strictly word-like by modulating nonwords' sub-lexical cues, which are properties related to the phonological constraints of a target language (e.g., length, phonotactic probability and prosody; Armon-Lotem, de Jong, & Meir, 2015; Thordardottir, 2017). Non language-like items are often developed with the aim of reducing the word-likeness of the stimuli and at the same time respecting the phonological constraints of a target language (e.g., non Icelandic-like stimuli for bilingual Icelandic-speaking children should be developed reducing the degree of Icelandic word-likeness and respecting Icelandic phonological constraints; Thordardottir, 2008). Finally, cross-linguistic nonwords follow and expand the principles of non language-like nonwords. They are developed according to the shared phonological constraints of a set of languages. Following this design, stimuli should be minimally sensitive to a bilingual child's language exposure and proficiency in a target language (Chiat, & Polišenská, 2016). Even though nonwords are often classified into two different categories (namely, language-like and non language-like), the manipulation of sub-lexical cues allows to develop nonwords as more language-like or less language-like (Szewczyk, Marecka, Chiat, & Wodniecka, 2018), suggesting the difference between these categories is not dichotomic but rather a continuum. To give some examples of the different types of nonwords, /ˈnɑskət/ is an English-like nonword (Chiat, & Polišenská, 2016), /vopekɛt/ is a Dutch-like nonword (Boerma, Chiat, Leseman, Timmermeister, Wijnen, & Blom, 2015), and /kata'sepo/ is an Italian-like nonword (Dispaldro, Leonard, & Deevy, 2013). Looking at non language-like items, /jolla/ and /vopgem/ are non Icelandic-like nonwords (Thordardottir, 2008), and /ˈsi ˈpulɑ/ (Chiat, & Polišenská, 2016) and /lʊmikɑ/ (Boerma et al., 2015) are cross-linguistic stimuli.

Additionally, the NWR presentation by examiners might influence children's NWR performance. In detail, language-specific prosodic and articulatory features can influence the degree of language-likeness of nonwords to a target language. As a consequence, theoretical frameworks bearing on the development of stimuli which are minimally language-specific should administer stimuli with the aim of maximally reducing language-specific suprasegmental features (Thordardottir, 2008). In the literature, stimuli have been audio-recorded prior to NWR administration and presented through digital devices (de Almeida, Ferré, Morin, Prévost, dos Santos, Tuller, Zebib, & Barthez, 2017; Summers, Bohman, Gillam, Peña, & Bedore, 2010), or presented orally (Kehoe, Poulin-Dubois, & Friend, 2021; Vaahtoranta, Suggate, Lenhart, & Lenhard, 2021).

Two notation systems have often been used in the literature to calculate NWR performance. Whole-word scoring measures the accuracy of each item as a whole – that is, each nonword repeated is scored as correct or incorrect and the number of nonwords correctly repeated is obtained. Phoneme scoring is based on the number of phonemes correctly repeated throughout the task, irrespective of accuracy in whole-word repetition. Different authors have tested the validity of the two scoring systems and similar results have been found overall in identifying monolingual children with atypical language development (Dispaldro et al., 2013; Graf Estes et al., 2007). Looking at bilingual children, while mixed evidence has been found favoring whole-word (Boerma et al., 2015) or phoneme (Guiberson, & Rodríguez, 2015) scoring as the better system for the identification of atypical language development, a recent meta-analysis underlined the absence of differences across the two scoring methods in diagnostic accuracy (Ortiz, 2021). Finally, similar results have been found across the two scoring methods in terms of correlations between NWR and other language measures (Brandeker, & Thordardottir, 2015).

Nonword repetition in bilingual children

NWR tasks have been found to identify language-learning difficulties in bilingual populations (Ortiz, 2021) and distinguish such difficulties from those due to different amount and degree of language-specific exposure and experience (Chiat, & Polišenská, 2016).

Different NWR types have been used with bilingual populations. Beyond the common use of language-like or non language-like stimuli, a recent research network attempted to develop language assessment tools which are minimally influenced by prior language-specific exposure (which is the time of exposure received by children in each of the languages they are exposed to before the assessment time) and proficiency, in order to maximally assess language-learning processes in bilingual children (Armon-Lotem et al., 2015). Following this theoretical framework, that same research network developed cross-linguistic (also called quasi-universal; Chiat, 2015) nonwords. In detail, cross-linguistic stimuli are developed respecting phonological constraints of a set of languages, e.g., 2- to 5-syllables long, with the CV syllabic structure, including consonants and vowels which are common among the set of languages, with prosody tuned with the target languages spoken by the child (Chiat, & Polišenská, 2016).

Different authors raised debates on the impact of prior language exposure and experience on NWR performance. Different scholars argued that prior language exposure to the languages spoken by the child might be related to NWR accuracy (Gibson, Summers, Peña, Bedore, Gillam, & Bohman, 2015; Schraeyen, Elst, Geudens, Ghesquière, & Sandra, 2018). Mixed evidence has been found on the association between language exposure to a specific target language and NWR performance (Bonifacci, Barbieri, Tomassini, & Roch, 2018; Core, Chaturvedy, & Martinez-Nadramia, 2017).

Nonword repetition performance and language exposure

Evidence supports the claim that prior exposure to a specific language partially explains heterogeneity in NWR performance (Antonijevic, Lyons, Malley, Meir, Haman, Banasik, Carroll, McMenamin, Rodden, & Fitzmaurice, 2019; Thordardottir, 2017), but mixed evidence has been found on the relationship between NWR and language exposure (Antonijevic et al., 2019; Barbosa et al., 2017; Engel de Abreu, Baldassi, Puglisi, & Befi-Lopes, 2013; Tuller, Hamann, Chilla, Ferré, Morin, Prevost, Santos, Ibrahim, & Zebib, 2018). Different hypotheses support either the independence or the relationship between NWR and language exposure.

On the one hand, since in NWR tasks children have to repeat items they have never heard before, it seems that prior language exposure should not, or at most minimally, affect NWR performance. Following this perspective and given that language exposure is related to language development, NWR might be used to analyze language development as a means of optimally reducing the influence of language exposure on task performance. As a consequence, low NWR scores might be mainly related to language processing abilities rather than to prior language exposure. For this reason, different authors argue that this task might be helpful in identifying atypical language development in bilingual children and in disentangling low language assessment scores due to lower exposure to the language of assessment from those due to language difficulties (Chiat, & Polišenská, 2016).

https://doi.org/10.1017/S1366728922000906 Published online by Cambridge University Press

On the other hand, NWR is often considered as a neuropsychological task that mainly assesses phonological short-term memory (Baddeley, 1986; Chiat, 2015). Indeed, the task requires storing, retrieving and reproducing a meaningless sequence of phonemes. Even though NWR is not a linguistic task that mainly assesses a language domain, it does involve phonological abilities, which is why prior language exposure should have at least a low impact on NWR. The impact of prior language exposure on NWR might be mediated by the nonwords' sub-lexical cues, which if designed accordingly can make nonwords more language-like and, therefore, enhance NWR performance. Mixed evidence has been found on the relationship between non language-like NWR performance and language exposure (Öberg, 2020; Vaahtoranta et al., 2021).

Different language exposure measures are frequently used to study its association with NWR – namely, age of first exposure, current exposure, and cumulative exposure. Age of first exposure is the chronological age at which the child has first been exposed to a specific language. Current exposure is the amount of exposure to a language calculated over a short period just before the assessment. Cumulative exposure is measured in diverse ways across studies. One definition is based on the amount of exposure to a language calculated in daily waking hours (e.g., Parra, Hoff, & Core, 2011); another definition relies on an index considering settings, speakers, and speakers' speech features (Thordardottir, & Brandeker, 2013), as well as the length of exposure to the major language at educational institutions (Duncan, & Paradis, 2016; Thordardottir, & Juliusdottir, 2013), across the child's lifetime.

When studying the association between NWR performance and language exposure, the latter is often calculated for the language according to which the nonwords have been developed – that is, the language from which the phonological constraints to develop the nonwords have been taken (e.g., if the stimuli are developed following English phonological constraints, language exposure is calculated for English; (Buac, Gross, & Kaushanskaya, 2016; Talli, & Stavrakaki, 2020); however, this is not always the case (Gathercole, & Masoura, 2005; Pérez-Navarro, Molinaro, & Lallier, 2020; Summers et al., 2010).

Potential moderators can affect the relationship between NWR and language exposure (Armon-Lotem, & Meir, 2016; Gutiérrez-Clellen, & Simon-Cereijido, 2010). Looking at nonwords' features, the nature of NWR stimuli (e.g., cross-linguistic vs language-specific) might have an influence on NWR performance in bilinguals (Chiat, & Polišenská, 2016). Similarly, the number of stimuli included in NWR lists varies across studies (Archibald, & Gathercole, 2006) and might also be related to NWR performance. Indeed, results from tasks with a lower number of nonwords might be less reliable because they provide less data on different features of nonwords (e.g., presence or absence of clusters, phonotactic probability, length variability). At the same time, lists with a higher number of stimuli might be affected by fatigue or attention decrease, which might impact negatively on NWR performance.

In addition, language exposure might be differently related to NWR performance in children with and without atypical language development. It might be the case that NWR performance in children with typical language development can be enhanced by language exposure, whereas children with difficulties in processing nonwords might be more influenced by language difficulties than prior language exposure (Boerma et al., 2015). At the same time, beyond language difficulties affecting NWR, prior language exposure might affect NWR performance in both children with and without atypical language development (de Almeida et al., 2017).

Finally, chronological age might also have a mediating effect on the interplay between children's NWR performance and their individual language exposure. In monolingual development, for instance, it has been proposed that, especially during the early stages of language development – up to the age of five years – mechanisms related to phonological short-term memory support language acquisition; while after this age the relation is inverted, with language knowledge supporting phonological memory (Coady, & Evans, 2008). In more detail, word learning and NWR tasks involve similar phonological short-term memory processes (e.g., to retain, store, retrieve, and eventually reproduce a meaningless or meaningful sequence of phonemes), all of which play a central role in carrying out both tasks successfully. It has further been suggested that after the first years of life children have mastered their language to a great extent. Not only have they acquired a relatively large lexicon but they have also developed a more comprehensive and reliable language knowledge across the various linguistic levels. Such broad language mastery could play a core role in children's NWR performance and might even outweigh the contribution of phonological short-term memory processes.

Following this perspective, we hypothesize that this same pattern might occur in bilingual language development too. Indeed, phonological short-term memory is involved in language acquisition whether a child is exposed to and learning one or more languages. In bilingual children, phonological short-term memory might be influenced by the relative amount of exposure a bilingual child receives in each of their languages. Since word learning and NWR tasks involve similar phonological short-term memory processes, such processes would play a core role in NWR accuracy during the first years of bilingual children's exposure to multiple languages, when they have just started acquiring their lexicon. Later on during language development, once bilingual children have received reliable exposure to each of their languages, their NWR performance will rely more on their knowledge of the languages than on language exposure. In particular, mastery of language-specific phonological constraints would enable language-specific NWR performance to be facilitated by sub-lexical cues. Given that, in general, older children have accumulated greater language exposure and, thus, greater language knowledge than younger children, who have had less language exposure and, thus, less language knowledge, it might be the case that language exposure and NWR are more strongly related in younger children than in older ones.

To the extent that NWR tasks are one of the most used assessment tools to identify language impairment (Schwob et al., 2021) in monolingual (Graf Estes et al., 2007) and bilingual (Ortiz, 2021) children, we wish to contribute to the better understanding of the impact of language exposure on NWR performance, as well as the contribution of child internal and external factors.

The current study

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

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

https://doi.org/10.1017/S1366728922000906 Published online by Cambridge University Press

Is the association between NWR and language exposure moderated 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 NWR scoring system (whole-word vs phoneme scoring)?
d) language development (typical or atypical)?
e) participants' chronological age (in toddlers, preschoolers and schoolers)?

Methods

Our research design follows PRISMA guidelines for systematic reviews and meta-analyses (Liberati, Altman, Tetzlaff, Mulrow, Gøtzsche, Ioannidis, Clarke, Devereaux, Kleijnen, & Moher, 2009). The current study was registered in the "International Prospective Register of Systematic Reviews" (PROSPERO; CRD42020173573). The first research question of the current systematic review studies the association between two variables, while the second one analyzes whether the meta-analysis main effect is influenced by the selected 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: ("nonword repetition" AND "language exposure" AND "bilingual"), while for ERIC we used [(non-word repetition OR nonword repetition OR pseudowords OR nwr) AND (language exposure OR input) AND (bilingual OR bilingualism OR multilingual OR multilingualism)]. We used two different sets of search keywords to adjust to the search settings of each database. Additionally, we collected research works through mailing lists and personal contacts. Searches on Google Scholar ended on 29/11/2021 and on ERIC on 31/11/2021. Study selection was carried out following these steps: abstract retrieval, abstract screening, full text retrieval and full text screening (see Table S1 in Supplementary Material for the literature screening). Literature screening at abstract level was performed on both databases by the first author, and by the second author on ERIC and a portion of the search results from Google Scholar. All search results judged as relevant by one or both reviewers were screened by both at full text level. Once the literature search and screening were finished, the first and second authors compared their inclusion decisions and, when needed, reached an agreement through a consensus process; when consensus could not be reached, the last author's advice was sought to make a decision. We finally calculated inter-rater reliability on literature screening before comparing extracted data across coders (Orwin, 1994). Inter-rater reliability was carried out on a portion of the screened literature.

Inclusion and exclusion criteria

To be included in the current systematic review, experimental data had to be related to bilingual children – that is, children below the age of 18 years who are exposed to at least two languages during their lifespan. We included studies that employed NWR tasks (but not word learning or sentence imitation assessment tools) and reported statistical results for the association between NWR and language exposure measures. We excluded effect sizes on the association between performance in NWR tasks developed following the phonological constraints of one language (e.g., Spanish) and language exposure measured for a different language (e.g., English) (Gathercole, & Masoura, 2005; Parra et al., 2011; Pérez-Navarro et al., 2020; Summers et al., 2010). Multiple comparison results not reporting the single effect of language exposure on NWR scores were excluded as well. When we found included studies from the same laboratories, we asked for further information about studies' samples to avoid duplicate data (from these procedures, we excluded Santos, & Ferré, 2016).

Data extraction

After consensus was reached regarding included studies, the first two authors carried out blind data extraction, which they then compared and discussed. For included articles, we extracted data about participants, NWR research design, language exposure measures, and statistical analyses. Regarding participants, we reported the sample size, the chronological age, and whether children had typical or atypical language development or if both were included. Regarding the NWR tasks, we extracted the type of nonwords used (language-like, non language-like or mixed), the amount of nonwords in the NWR lists and their syllabic range (i.e., their different lengths, in number of syllables), and the scoring system employed (phoneme and/or whole-word scoring). For language exposure we extracted the measure used (e.g., cumulative exposure, current exposure, age of first exposure). We reported the statistical analysis, statistical results, and the related significance (see Table S2 in Supplementary Material for the data extraction). Finally, we analyzed the agreement on data extraction across coders using percentage.

Risk of bias

We developed a list of study risk-of-bias variables after consulting the 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 gray literature; (c) bilingual status across participants (if they share only the major language or also the native one); (d) parents' bilingual status (both or only one of the parents shares with the child a native language different from the major language of a geographical area); (e) data bearing on participants' native and/or major language (language exposure and nonwords based on and developed following the native and/ or major language); (f) amount of nonwords administered (between 16 and 24 stimuli; between 8 and 16; between 24 and 40 or more than 40 stimuli); (g) the nonwords' syllabic range (three groups classified for range and maximum length: the first group included lists with a syllabic range as [1-5], [1-4] or [2-5]; the second group included lists with nonwords of two or three different lengths excluding 6-syllable stimuli and above; the third group included lists having either all nonwords of the same length or nonwords of more than four different lengths as well as lists including 6-syllable items and above; the latter were grouped together in the same subgroup because they included more extreme syllabic ranges which could potentially lead to more extreme scores). We did not exclude studies based on the risk of bias. Rather, we took an inclusive approach to study selection, to maximize the literature of our systematic review and meta-analysis, and we collected data related to the risk of bias.

https://doi.org/10.1017/S1366728922000906 Published online by Cambridge University Press

A checklist of desirable study characteristics is given in Table S3 in Supplementary Material. For each study we assessed whether each of the desirable study characteristics was present, relatively present, or absent.

Data processing

We coded both correlation and comparison results and then converted them into Fisher's z scores. While correlation coefficients were directly converted into z scores, we followed Lakens (2013) to calculate Fisher's z score from analysis of variance. We decided not to reproduce those statistical analyses already reported in the publications, nor additional ones, on the raw descriptive statistics of the included studies.

Meta-analysis

We conducted a random-effects model meta-analysis for each individual predictor using Review Manager 5.4 (RevMan; The Cochrane Collaboration, 2020). We acknowledged significance at $p < .05$.

Main effect

We reported for each study standard errors (SE) and Fisher's z scores as effect sizes. The meta-analysis main effect is calculated and reported with odd ratios by RevMan. The subgroup analysis results are calculated following the same statistical procedure.

Heterogeneity

We estimated the magnitude of heterogeneity using the I^2 value (Borenstein, Hedges, Higgins, & Rothstein, 2009). The interpretation of I^2 was insignificant heterogeneity for 0%–25%, low heterogeneity for 26%–50%, moderate heterogeneity for 51%–75%, and high heterogeneity for > 75% (Higgins, Thompson, Deeks, & Altman, 2003).

Moderation analysis

We tested potential moderators of the relationship between predictors and outcomes using subgroup analyses. Recommendations indicate that moderation analyses are appropriate when there is at least low heterogeneity ($I^2 > .25$) on the main effect and a minimum of eight studies for each subgroup (Borenstein et al., 2009).

In all subgroup analyses, we tested the moderators that we hypothesized were relevant to one specific predictor. Concerning language exposure, we analyzed whether the main effect differed according to the language exposure measure (cumulative exposure, current exposure, or age of first exposure), the NWR stimuli type (language-like vs non language-like; we expected the main effect to be stable when using language-like stimuli but maybe weaker when using non language-like stimuli, because the latter are developed with the aim to be minimally affected by language exposure), the NWR scoring system (phoneme vs whole-word scoring), the language development (typical or atypical), and the participants' chronological age (in toddlers [0–3 years], preschoolers [3–6 years], and scholars [older than 6]).

Eight subgroups did not conform to the recommendations on moderation analysis because they included less than eight studies: NWR and language exposure using non language-like nonwords, in children with atypical language development, in toddlers, in children younger than six, in children sharing their native language with at least one parent, with data bearing on native language, for NWR lists containing 8–16 or 24–40 stimuli and for NWR lists with a single length, with more than four different lengths, or including 6-syllable items and above. Subgroup analysis was not carried out for these subgroups; however, we do report the single effect sizes of each study in the Results section in order to open qualitative interpretations to such analyses.

Publication bias and sensitivity analysis

We carried out publication bias and sensitivity analyses to evaluate the validity and robustness of the meta-analysis findings. We assessed publication bias by examining funnel plots for asymmetry, as well as conducting subgroup analyses (see Table S3 in Supplementary Materials for further information). We assessed sensitivity by exploring the effects of removing each individual study on our meta-analysis main effect and on each subgroup analysis (Fisher, 2017). We only report sensitivity results that change the main effect. Sensitivity analyses were conducted on subgroups with more than eight studies.

Results

Systematic review

See Appendix S1 in Supplementary Material for the PRISMA flow diagram detailing search results and records excluded for various reasons. Among 882 screened research articles, a total of 24 met our selection criteria and were included (see Table S1). We also analyzed inter-rater reliability on literature screening and agreement on data extraction of included studies. We found almost perfect agreement (McHugh, 2012) on literature screening (Cohen's κ=.967; see Appendix S2 for reproducible data) and an agreement of 94.55% in data extraction (see Appendix S3 for reproducible data). Out of the 24 included articles (see Table 1 for the main characteristics of the included studies), 16 have been published in peer-reviewed journals as experimental studies. Among the remaining eight works, there was a study under submission (Pérez-Navarro et al., 2020), a study published in conference proceedings (Core et al., 2017), three PhD dissertations (Huls, 2017; Kołak, 2020; Öberg, 2020), and three master theses (Li'el, 2017; Limacher, 2019; Reid, 2019).

Sample sizes were heterogeneous across studies, ranging from 16 to 151 children (M = 58.3, Mdn = 55.5) and with a total of 1399 children. We found 12 studies including more than 50 participants. Participants were recruited from toddlerhood to high school, with mean age ranging from 22 to 134 months (M = 69.9, Mdn = 69.75). Eighteen studies included only children with typical language development, one study reported data only on children with atypical language development (Vaahtoranta et al., 2021) and five reported data on both children with typical and atypical development. Among these five works, two reported separate effect sizes from subgroups of children with typical and atypical language development.

Looking at participants' linguistic backgrounds, 14 studies reported data on samples composed of bilingual children exposed to the same set of languages. Thirteen studies reported English as a major language and Spanish, Chinese or other South Asian languages, French, Polish, or Welsh as the participants' native language, while one study reported data on bilingual children learning English at school with Greek as their native language. Others included children having French, English, Hebrew, Italian, Icelandic or Australian English as their major language and a range of different native languages. One study included Basque-Spanish bilingual children. One study enrolled bilingual children in Luxembourg, a trilingual country where French and

https://doi.org/10.1017/S1366728922000906 Published online by Cambridge University Press

https://doi.org/10.1017/S1366728922000906 Published online by Cambridge University Press

Table 1. Participant and methodological characteristics of included records and unpublished reports.

Study (n=24)	Publication type	N	Age in months M (SD)[1]	Languages spoken by children	NWR type	NWR scoring	Language exposure measure	Statistical analysis
Altman, Feldman, Yitzhaki, Lotem, & Walters, 2014	international journal article	65	72 [48.72-72.11]	Russian-Hebrew	language-like (42 [1-4]-syllable Russian- and Hebrew-like stimuli)	whole-word	cumulative exposure	regression (single effect of cumulative exposure on NWR from two-way ANOVA)
Antonijevic et al., 2019	international journal article	88	80.4 (6.89)	NLs[2]-Irish [English]	both non language-like (16 [2-5]-syllable NWR) and language-like (24 [2-4]-syllable Irish English-like)	whole-word	age of first exposure, monolingual vs bilingual NWR accuracy	Pearson's correlation, *t*-test
Brandeker & Thordardottir, 2015	international journal article	48	33.67 (3.92)	French-English	language-like (15 [1-3]-syllable French-like NWR)	both whole-word and phoneme	cumulative exposure	correlation
Core, Chaturvedi & Martinez-Nadramia, 2017	conference proceedings	105	30.50 (na)	Spanish-English	language-like (12 English-like and 12 Spanish-like [2-4]-syllable nonwords)	consonant	cumulative exposure	correlation
de Almeida et al., 2017	international journal article	82	76.27 (na)	(Arabic, Portuguese or Turkish)-French	both non language-like (30 [1-3]-syllable LITMUS-FRA-NWR) and language-like (41 [1-3]-syllable French-like)	whole-word	use of French at home and with friends, age of first exposure, length of exposure, language dominance index	Spearman's rho correlation
Duncan & Paradis, 2016	international journal article	75	69.60 (5.67)	(Chinese language or South Asian language)-English	language-like (18 [1-7]-syllable English-like nonwords; CTOPP)	whole-word	cumulative exposure	stepwise regression model
Engel de Abreu et al., 2013	international journal article	40	85.00 (3.2)	Portuguese-Luxembourgish	language-like (40 [2-5]-syllable Brazilian- and Portuguese-like NWR)	whole-word	bilingual vs bilingual NWR accuracy on bilingual groups with different exposure to dominant language (Luxembourgish)	*t*-test, Cohen's *d*
Farabolini et al., 2021	international journal article	19	51.35 (5.5)	NLs-Italian	language-like (16 [2-4]-syllable Italian-like)	both phoneme and whole-word	cumulative exposure, current exposure, and age of first exposure	Spearman's rho correlation
Gathercole & Masoura, 2005	international journal article	80	134 (15)	Greek-English	language-like (50 [2-5]-syllable English-like; CNRep list)	whole-word	cumulative exposure (English tuition attendance at school)	correlation
Gibson et al., 2015	international journal article	52	69.9 (5.28)	Spanish-English	language-like (12 Spanish- and 12 English-like [2-4]-syllable NWR)	phoneme	cumulative exposure	ANOVA
Huls, 2017	master thesis	47	71 (6)	Spanish-English	both language-like (16 [2-5]-syllable Spanish- and English-like items) and non language-like (16 [2-5]-syllable stimuli)	phoneme	age of first exposure, current English exposure and use (after controlling for current English use)	multiple linear regressions

Kehoe, Poulin-Dubois & Friend, 2021	international journal article	48	[35.2-56.2]	(Spanish or French)-English	language-like (40 [1-5]-syllable English-like NWR)	phoneme	current exposure	correlation coefficent
Kołak, 2020	PhD dissertation	28	29.04 (4.2)	Polish-(Scottish-Irish-British) English	language-like (50 [2-4]-syllable English-like nonwords)	whole-word	age of first exposure	Sperman's rho correlation
Li'el, 2017	master thesis	61	69.65 (7.55)	NLs- Australian English	language-like (18 [1-7]-syllable English-like nonwords; CTOPP)	whole-word	age of first exposure, length of exposure	correlation
Limacher, 2019	master thesis	34	57.39 (10.01)	NLs-English	language-like (18 1-to-7-syllable English-like NWR; CTOPP list)	whole-word	monolingual vs bilingual NWR accuracy, age of first exposure and length of exposure to English	*t*-test, correlation
Öberg, 2020	PhD dissertation	98	73 (na)	Arabic-Swedish	both non language-like (16 [2-5]-syllable NWR) and language-like (24 [2-5]-syllable NWR)	whole-word	current exposure and length of exposure	Pearson's correlation coefficent
Parra, Hoff & Core, 2011	international journal article	41	22.78 (0.39)	Spanish-English	language-like (12 [1-3]-syllable Spanish-like)	phoneme	current English exposure	correlation
Pérez-Navarro, Molinaro, & Lallier, 2020	unpublished data	63	51.61 (1.12)	Basque-Spanish	language-like (24 [2-5]-syllable stimuli)	whole-word	age of first exposure to Basque	Pearson's correlation
Reid, 2019	master thesis	28	117.2 (17.95)	NLs-French	language-like ([na-8]-syllable French-like nonwords)	phoneme	cumulative exposure	correlation
Sharp & Gathercole, 2013	international journal article	45	66.4 (na)	Welsh-English	language-like (108 [1-2]-syllable Welsh-like nonwords)	phoneme	current exposure	regression
Summers et al., 2010	international journal article	59	66.60 (na)	Spanish-English	language-like (17 [2-4]-syllable Spanish-like)	phoneme	age of first exposure, current exposure, children's language use	regression model
Thordardottir & Juliusdottir, 2013	international journal article	16	134.10 (15.7)	NLs-Icelandic	both language-like and non language-like (25 [1-5]-syllable Icelandic-like and 25 [1-5]-syllable non language-like)	phoneme	age of arrival	*t*-test
Tuller et al., 2018	international journal article	151	83.97 (13.3)	(Arabic, Portuguese or Turkish)-French	both non language-like (30 [1-3]-syllable LITMUS NWR) and language-like (36 [1-3]-syllable German- or French-like nonwords)	whole-word	early NL exposure, current L2[3] richness	correlation
Vaahtoranta et al., 2021	international journal article	65	63.96 (9)	NLs-German	Both non language-like (16 stimuli) and language-like (18 German-like nonwords)	whole-word	length of German exposure, current exposure to German and the NL	Spearman's rho correlation

[1]When mean and standard deviation were not available, age range is included in square brackets.
[2]NL = native language. NLs indicates bilingual participants in the sample had different native languages.
[3]L2 = dominant language in a specific geographical area.

https://doi.org/10.1017/S1366728922000906 Published online by Cambridge University Press

German are learned at school; these children had either Portuguese or Brazilian Portuguese as their native language, and Luxembourgish as their major language (see Table 1 for further details).

Different language exposure measures have been used by researchers. The most common were cumulative exposure (N = 9), current exposure (N = 9), and age of first exposure (N = 9). The latter measure included age of arrival at the geographical area where the target language is spoken and age of first contact with the target language.

Included studies reported effect sizes using only language-like (N = 18) or both language-like and non language-like nonwords (N = 6). Four 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). Among the six studies using both non language-like and language-like stimuli, three reported effect sizes for each NWR type separately. Among the 18 studies including language-like stimuli, two of them reported associations between exposure to the participants' native language and NWR with nonwords developed according to the phonological constraints of the participants' native language. Thirteen other effect sizes corresponded to associations between exposure to the major language and NWR with nonwords developed according to the phonological constraints of the major language. Two studies used stimuli developed following the phonological constraints of both the native and the major language of the participants. Additionally, one study used language-like stimuli constructed according to the phonological constraints of the standard variety of a language with children exposed to a geographical variety of that language (North American English-like nonwords used for bilingual children having Australian English as their major language).

Eight studies used an amount of nonwords ranging between 16 and 24, and five studies used either less than 16 or between 24 and 40; 10 studies used more than 40 stimuli, and one study did not report this information. Looking at the syllabic range of the NWR lists used by each study, eight studies had a syllabic range of 1–5 (i.e., the shortest nonwords were monosyllabic and the longest ones had five syllables), 1–4, or 2–5. These three syllabic ranges can be considered as the most optimally reliable since they include a wide range of word length that is common across languages. Thirteen studies used stimuli with two or three different lengths that varied from one study to another but in all cases excluded 6-syllable nonwords and above. The remaining three works used NWR lists with more than four different lengths, or including nonwords with six or more syllables. Looking at NWR presentation procedures, 12 studies presented audio-recorded stimuli through digital devices, six administered stimuli orally and six did not report this information. Finally, 13 research works used whole-word scoring, nine used phoneme scoring, while two works calculated NWR accuracy using both whole-word and phoneme scoring and reported separate results from each notation system.

Meta-analysis

Quantitative integration

We carried out a random-effects meta-analysis on the selected studies (see Appendix S4 in Supplementary Material for reproducible data). Across the 24 studies, the main effect revealed a positive and significant association between prior language exposure and NWR performance (*OR*: 1.15 [1.06, 1.25], $p < .0005$). Figure 1 displays the forest plot for the 24 included studies. The funnel plot in Figure 2 shows low heterogeneity across studies ($I^2 = 44\%$). Sensitivity analysis on the included studies revealed

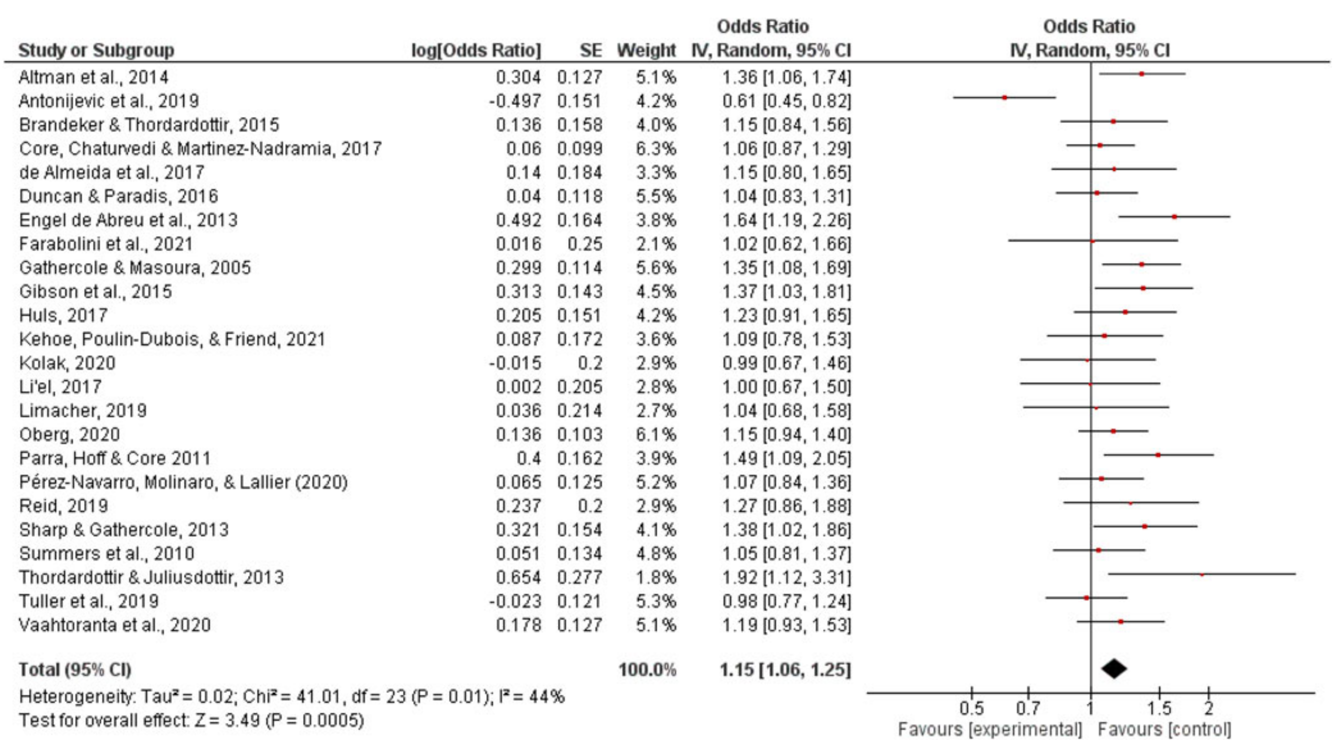

Figure 1. Forest plot of the association between nonword repetition and language exposure.

https://doi.org/10.1017/S1366728922000906 Published online by Cambridge University Press

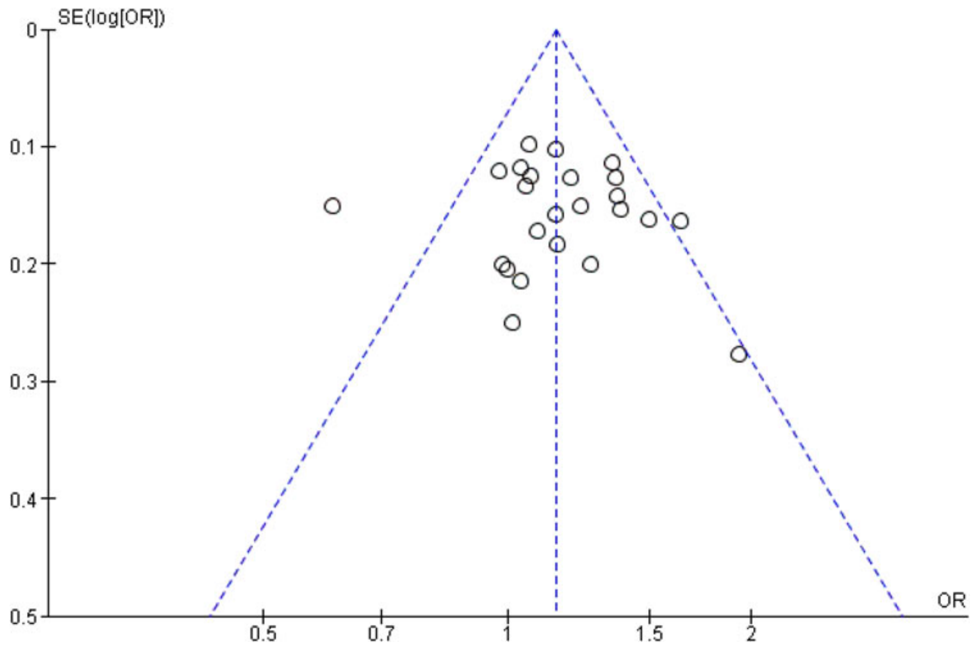

Figure 2. Funnel plot of the association between nonword repetition and language exposure.

the main effect does not change when removing each included study separately.

We then carried out subgroup analyses to study the effect of different variables that can affect the main effect (see Table 2 for summary findings and Appendix S5 for forest and funnel plots of each subgroup analysis).

Regarding the impact of the language exposure measure, while better performance on NWR measures was exhibited by children with higher levels of cumulative (*OR*: 1.18 [1.07, 1.29], $p = .00007$, $I^2 = 4\%$) or current (*OR*: 1.26 [1.15, 1.39], $p < .00001$, $I^2 = 0\%$) exposure, the same was not true for age of first exposure, which did not seem to be related with NWR performance (*OR*: 0.98 [0.80, 1.18], $p = .81$, $I^2 = 58\%$).

Concerning NWR research designs, the main effect did not change when using language-like items (*OR*: 1.15 [1.05, 1.25], $p = .002$, $I^2 = 46\%$).

Looking at the NWR scoring method, the association between NWR and language exposure remained significant when using phoneme scoring (*OR*: 1.17 [1.02, 1.34], $p = .02$; $I^2 = 60\%$), although the sensitivity analysis revealed that the main effect changed in one case (*OR*: 1.15 [1.00, 1.32], $p = .06$; $I^2 = 60\%$ when removing Parra et al., 2011). When using whole-word scoring, the main effect approached but did not reach significance (*OR*: 1.11 [1.00, 1.24], $p = .05$; $I^2 = 54\%$); sensitivity analysis revealed that the main effect was still significant in three cases (*OR*: 1.11 [1.00, 1.24], $p < .00001$; $I^2 = 0\%$ when removing Antonijevic et al., 2019; *OR*: 1.12 [1.00, 1.25], $p = .04$; $I^2 = 56\%$ when removing Kołak, 2020; *OR*: 1.12 [1.01, 1.26], $p = .04$; $I^2 = 57\%$ when removing Tuller et al., 2018). Then, the main effect did not change when considering effect size only from children with typical language development (*OR*: 1.17 [1.10, 1.25], $p = .00001$, $I^2 = 14\%$).

Considering chronological age, when looking at effect sizes collected in schoolers and beyond, the association between NWR and language exposure did not reach significance (*OR*: 1.17 [0.98, 1.38], $p = .08$, $I^2 = 68\%$); sensitivity analysis revealed that the main effect was still significant in one case (*OR*: 1.24 [1.11, 1.39]; $p = .0001$; $I^2 = 18\%$ when removing Antonijevic et al., 2019).

Finally, we looked at the effect of selection bias on the meta-analysis (see Appendix S6 for further information).

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

Sub-group analysis	N of studies	N of participants	Results (OR)	Significance	I^2(%)
NWR & cumulative exposure	9	464	1.18 [1.07, 1.29]	.0007	4
NWR & current exposure	9	451	1.26 [1.15, 1.39]	<.00001	0
NWR & age of first exposure	9	376	0.98 [0.80, 1.18]	.81	58
Language-like NWR & language exposure	21	1133	1.15 [1.05, 1.25]	.002	46
NWR & language exposure using whole-word scoring	15	948	1.11 [1.00, 1.24]	.05	54
NWR & language exposure using phoneme-scoring	13	615	1.17 [1.02, 1.34]	.02	60
NWR & language exposure in samples with typical lang. dev.	20	1053	1.17 [1.10, 1.25]	<.00001	14
NWR & language exposure in schoolers	11	663	1.17 [0.98, 1.38]	.08	68

Concerning the representativeness of the cohort, the main effect of the meta-analysis did not change considering studies with a sample size larger than 50 participants (*OR*: 1.14 [1.06, 1.22], $p = .0003$, $I^2 = 0\%$). When considering studies with a sample size smaller than 50, the main effect approached and did not reach significance (*OR*: 1.18 [0.99, 1.41]; $p = .06$; $I^2 = 65\%$), but sensitivity analysis revealed the main effect did not change in one case (*OR*: 1.27 [1.14, 1.42]; $p < .0001$; $I^2 = 1\%$ when removing Antonijevic et al., 2019).

Then, considering the publication status of the work, while the main effect did not change when considering only evidence from published articles (*OR*: 1.17 [1.05, 1.31], $p = .004$, $I^2 = 59\%$), it approached but did not reach significance when considering only gray literature (*OR*: 1.11 [1.00, 1.22], $p = .05$, $I^2 = 0\%$).

Looking at bilingual status between participants, the main effect was still significant for participants sharing both major and native languages (*OR*: 1.21 [1.13, 1.31], $p < .00001$, $I^2 = 6\%$). For participants sharing only the major language, the association was not significant (*OR*: 1.11 [0.89, 1.38], $p = .36$, $I^2 = 71\%$), but sensitivity analysis revealed the main effect did not change in one case (*OR*: 1.20 [1.02, 1.41], $p = .03$, $I^2 = 35\%$ when removing Antonijevic et al., 2019). Concerning research designs using language-like NWR, the main effect was still significant across data bearing on major languages (*OR*: 1.18 [1.09, 1.28], $p < .0001$, $I^2 = 22\%$).

Finally, regarding the length of NWR lists, while the main effect was still significant for lists with more than 40 nonwords (*OR*: 1.25 [1.12, 1.41], $p = .0001$, $I^2 = 30$), it was no longer significant for 16–24 nonword lists (*OR*: 1.01 [0.86, 1.19], $p = .091$, $I^2 = 57\%$).

Last, across syllabic range, the main effect did not change for the 1–4, 1–5, or 2–5 syllabic range group (*OR*: 1.27 [1.13, 1.42], $p < .00001$, $I^2 = 31\%$). Similar results have been found for nonwords of two or three different lengths, excluding 6-syllable nonwords and above (*OR*: 1.11 [1.01, 1.22], $p = .03$, $I^2 = 24\%$), but sensitivity analysis revealed the main effect changed in four cases (*OR*: 1.09 [0.99, 1.19], $p = .09$, $I^2 = 17\%$ when removing Gibson et al., 2015; *OR*: 1.10 [1.00, 1.22], $p = .06$, $I^2 = 27\%$ when removing Kehoe et al., 2021; *OR*: 1.08 [0.99, 1.18], $p = .08$, $I^2 = 8\%$ when removing Parra et al., 2011; *OR*: 1.09 [0.99, 1.20], $p = .08$, $I^2 = 18\%$ when removing Sharp, & Gathercole, 2013).

Qualitative analysis

Even though subgroup analyses were not conducted for subgroups with fewer than eight studies, here we report the individual results of those studies with the aim of sharing evidence for qualitative analysis.

Three included studies reported data using non language-like stimuli. In detail, they all used cross-linguistic items. Among these three studies, two reported a non-significant association between NWR performance and language exposure (Huls, 2017; Vaahtoranta et al., 2021), while one article reported mixed evidence (Öberg, 2020). Two other included studies used cross-linguistic nonwords but they report overall NWR performance including both language-like and cross-linguistic nonwords (de Almeida et al., 2017; Tuller et al., 2018).

Taking into account studies on children with atypical language development only, three of them found a non-significant association between NWR and language exposure (de Almeida et al., 2017; Li'el, 2017; Vaahtoranta et al., 2021).

Looking at the moderator effect of subgroups with children younger than six years, mixed evidence has been found. In detail, results collected from studies on toddlerhood revealed significant and positive associations (Core et al., 2017; Parra et al., 2011), significant and negative associations (Core et al., 2017), and non-significant associations (Kehoe et al., 2021), while one study found both significant and nonsignificant associations (Kołak, 2020). Data collected on preschoolers revealed non-significant associations between NWR and language exposure (Farabolini et al., 2021; Kehoe et al., 2021; Limacher, 2019; Pérez-Navarro et al., 2020; Sharp, & Gathercole, 2013), except for one study that found both significant and nonsignificant associations (Vaahtoranta et al., 2021).

Among the five studies with data bearing on native language, non-significant (Core et al., 2017; Pérez-Navarro et al., 2020; Vaahtoranta et al., 2021), significant (Engel de Abreu et al., 2013), and mixed results (Sharp, & Gathercole, 2013) have been found. Considering parents' bilingual status, no study included children sharing their native language (which is not the major language of the geographical area) with both parents, while on samples with at least one parent speaking the native language, significant (Duncan, & Paradis, 2016; Engel de Abreu et al., 2013; Parra et al., 2011), non-significant (Farabolini et al., 2021; Li'el, 2017; Limacher, 2019) and mixed (Kołak, 2020) evidence has been found.

Similarly, for NWR lists containing 8–16 or 24–40 stimuli, significant (Parra et al., 2011) and non-significant (Brandeker, & Thordardottir, 2015; Kehoe et al., 2021; Öberg, 2020; Vaahtoranta et al., 2021) associations between NWR and language exposure were found. Finally, looking at the nonwords' syllabic length, three studies used NWR lists with a single length, with more than four different lengths, or including 6-syllable items and above, and they all found non-significant associations between NWR and language exposure (Li'el, 2017; Limacher, 2019; Reid, 2019).

Discussion

Our meta-analysis investigated the association between prior language exposure and NWR performance in bilingual children, and we found a significant and positive correlation. We then carried out subgroup analyses to further examine which variables might affect the association. These revealed that the main effect remained significant when considering only studies that used cumulative exposure, current exposure, language-like stimuli, phoneme scoring, or data from children with typical language development. On the contrary, the association between NWR and language exposure was not significant when using age of first exposure or data from children older than 6, and it was weak when using whole-word scoring. The main effect was biased by the representativeness of the exposed cohort (i.e., the main effect changed when considering only studies with a sample size lower than 50), by participants' bilingual status (the main effect was no longer significant on samples composed by participants sharing only one language), and by publication status (it approached significance when considering only gray literature). Finally, the main effect was biased by the number of stimuli included in the NWR list (studies with 16-24 nonword lists changed the main effect).

We found that both current and cumulative exposure were associated with NWR performance, while, interestingly, age of first exposure to a language was not. A possible interpretation of this finding is that the former measures might be more representative of and more closely associated with language

https://doi.org/10.1017/S1366728922000906 Published online by Cambridge University Press

development in bilingual children than the latter. Another possible interpretation is that both current and cumulative exposure are measures focused on the amount of exposure received, while age of first exposure is related to when it is received, more particularly to its onset, and thus to considerations related to the sensitive period for language acquisition and development. Another possibility is that age of first exposure might play a core role in early language development (e.g., in the first three years of life), while the participants of the included studies are older (with a mean age of 68.4 months).

Nearly all the studies included in this meta-analysis employed quantitative language exposure measures. Quality of input features (e.g., exposure to native as opposed to non-native speakers, intra- and inter-language variability among speakers, speakers' lexicon, syntactic complexity, and variability in the use of concrete as opposed to abstract conversations) should be taken into account to analyze the moderator effect on the association between the role of language exposure and linguistic experience on language development (Anderson, Graham, Prime, Jenkins, & Madigan, 2021; Hoff, 2020). Unfortunately, few works measure and analyze the quality of input. Similarly, parental beliefs and expectations about language proficiency and the importance of each of the languages the child is exposed to might play a role in the weight with which language exposure can affect language development (Ronderos, Castilla-Earls, & Marissa Ramos, 2021).

Looking at the type of NWR stimuli, positive and significant associations were found for studies using language-like nonwords. These results support the assumption that language-like stimuli are related to language exposure in the target language. So, when using language-like stimuli on bilingual populations, the language-specific exposure received by a bilingual child on the target language is related to NWR performance. Moderation analysis on non language-like stimuli was not run since only three studies reported effect sizes using non language-like stimuli. These three studies all report non-significant associations between NWR and language exposure, which are in line with the idea that non language-like stimuli should maximally reduce the impact of exposure to a target language on NWR (Chiat, & Polišenská, 2016). However, this evidence should be taken with caution since it came from single effect sizes of three studies, not from moderation analysis of the current meta-analysis.

Looking at the effect of the NWR scoring system, we found the main effect does not change when using phoneme scoring while, surprisingly, it approaches but does not reach significance when using whole-word scoring (but sensitivity analysis revealed the main effect is still significant in three cases). This finding of a weak significance is not in line with previous results which suggested that both NWR scoring systems can be employed similarly to study the impact of language exposure on NWR performance (Brandeker, & Thordardottir, 2015; Farabolini et al., 2021). A possible interpretation is that while phoneme scoring analyzes the phonological processes involved in phonological short-term memory, which is associated with language exposure, it might be the case that whole-word scoring relies mainly on item-level processing, which might involve mechanisms closer to lexical processing and more independent from language exposure.

We found that the association between NWR and language exposure was not moderated by data from children with typical language development, while subgroup analysis on children with atypical language development could not be run due to the low number of studies in this subgroup. We encourage further research into this because it is possible that the variability introduced by the language difficulties of children with atypical language development is such that it overshadows other factors like language exposure. Indeed, not only is atypical language development an umbrella term covering various types of language difficulties for which different identification criteria are used in the literature, but these difficulties also present somewhat differently from one child to another. Hence, the impact of language exposure on NWR performance might be secondary, in the presence of atypical acquisition patterns, to underlying mechanisms related to language difficulties.

Finally, regarding chronological age, the main effect changed when considering evidence from schoolers older than six years. Following the hypothesis that NWR and word learning involve similar phonological short-term memory processes, at earlier stages of language acquisition, language exposure might enhance these mechanisms, which might in turn result in the association between language exposure and performance in NWR. Later on, since language exposure to a target language enhances phonological and lexical development in that language, older bilingual children might have reached both phonological and lexical abilities which are less dependent on language exposure: at this stage, NWR might be more related to such phonological and lexical abilities than to language exposure. Further studies should disentangle the contribution of language exposure and phonological short-term memory processes involved in tasks requiring retaining, storing, retrieving, and reproducing a linguistic sequence across age.

Regarding the risk of bias assessment, the main effect changed when considering studies with fewer than 50 children and studies published as gray literature. The lack of significance in studies with sample sizes lower than 50 might be due to a lack of statistical power. In turn, studies with low sample size cannot be considered representative of a target population. The weak significance of the association between NWR and language exposure for non peer-reviewed works might suggest there is a bias against the publication of negative results.

Many studies did not report the parents' bilingual status, so there was a lack of information regarding the subgroups we defined related to this information (both or at least one parent speak their native language to the child which is different from the major one). Thus, no interpretations can be elaborated on the moderation of the target variable. We hope further research can address this to explore the potential role of having one or both parents speaking one or more languages on children's NWR performance and, in general, on language development.

Then, the main effect changed when considering studies with bilingual children sharing the major language but having been exposed to different languages. This might be explained by the fact that bilingual populations with different native languages have acquired language through language-specific constraints. Additionally, recent research suggested that also cultural, sociolinguistic and pragmatic rules influence language development (Cristia, Farabolini, Scaff, Havron, & Stieglitz, 2020; Loukatou, Scaff, Demuth, Cristia, & Havron, 2021). From this perspective, we advance the hypothesis that such constraints and rules might also be related to NWR performance.

We underline the heterogeneity in both the number of nonwords used to calculate NWR performance (ranging from 12 in Parra et al., 2011 to 108 in Sharp, & Gathercole, 2013; mean = 36.72) and nonwords' syllabic range (ranging from syllabic ranges of 1–2 in Sharp, & Gathercole, 2013 to 1–7 in Duncan, & Paradis, 2016). The number of nonwords was a moderator of the association studied: the main effect changed for an amount of stimuli between 16 and 24 but did not when considering studies using more than 40 stimuli. This evidence does not support the hypothesis that a high amount of nonwords affects NWR due to, for example, decrease of attention or fatigue. At the same time, this result underlines that language exposure is associated with NWR lists including more than 40 stimuli; when using NWR lists with a high amount of stimuli, scholars and practitioners should bear in mind that NWR performance seems to be associated with language exposure.

Concerning syllabic range subgroup analyses, the association between language exposure and NWR lists with nonwords of two or three different lengths excluding 6-syllable nonwords and above is still significant, but sensitivity analysis revealed that the association is weak. The studies from these subgroups included syllabic ranges as 1-2, 1-3 or 2-4. One possible interpretation is that the lists containing stimuli with a narrow syllabic range might not be related to language exposure. An alternative hypothesis is that such syllabic ranges mainly involve phonological processes which are more related to mechanisms underlying phonological short-term memory than language exposure. In detail, it might be possible that these stimuli with relatively short length are mainly related to phonological processing of nonwords rather than to previous language exposure, but further research is needed. We highlight that we did not control for nonword length, and we suggest further studies should shed light on the impact of sub-lexical cues on the association between NWR and language exposure.

Finally, we also underline that different procedures have been found regarding NWR presentation, which are often administered orally or digitally through audio-recorded stimuli. The NWR presentation mode was not directly related with year of publication (e.g., Summers et al., 2010 used digital presentation; Kehoe et al., 2021 used oral presentation) nor with NWR type; further research should consider experimenters' prosodic and articulatory features which could influence NWR administration and, as a consequence, NWR performance. Audio-recorded administration might be the optimal methodological choice (Sahlen, Reuterskiold-Wagner, Nettelbladt, & Radeborget, 1999) to maximally reduce language-specific suprasegmental features for stimuli, as well as possibly ensuring homogeneity of administration to all participants.

Concerning clinical implications, we suggest, in line with previous literature, that NWR should be used together with other assessment tools, such as receptive and expressive lexical tasks (Haman, Wodniecka, Marecka, Szewczyk, Białecka-Pikul, Otwinowska, Mieszkowska, Łuniewska, Kołak, Miękisz, Kacprzak, Banasik,, & Foryś-Nogala, 2017), narrative tests (Gagarina, Klop, Kunnari, Tantele, Välimaa, Balčiūnienė, Bohnacker, & Walters, 2012), or sentence repetition lists (Meir, Walters, & Armon-Lotem, 2015), in order to have information from different language domains.

Limitations

Several limitations have to be reported regarding included articles in the current review. First, we excluded reports written in languages other than English, Spanish, French or Italian.

Second, there are differences across studies that might cause heterogeneity in our work. Sources of heterogeneity might be related to NWR features (e.g., type, scoring, administration, amount and length of stimuli) and bilingual background (e.g., exposure received, language exposure measure, parents' native languages, same or different languages spoken by the children, participants having the same or different native languages).

Third, there are research articles reporting quantitative results only for significant effect sizes (e.g., Kołak, 2020). Therefore, non-significant effect sizes from those studies could not be included in the meta-analysis. Such results, had they been included, possibly could have modified our results.

Fourth, risk of bias assessment is considered mandatory in meta-analysis considering randomized-trial designs (Boutron, Page, Higgins, Altman, Lundh, Hróbjartsson, & Group, 2019), which is not the case of our study. However, we decided to carry it out to obtain more information on the effect of potential bias on our main effect. It should be noted that the categorization of some of the selected biases (e.g., amount of nonwords and syllabic range) was chosen in an arbitrary fashion. Additionally, the impact of other sub-lexical cues (e.g., length, wordlikeness, phonotactic probability) on NWR was not analyzed in the current study and thus we cannot rule out the possible impact of these features on NWR accuracy. Mixed evidence has been reported in the literature on the effect of nonword length on NWR performance. While some authors found that NWR performance decreases as nonword syllabic length increases (Chiat, & Polišenská, 2016; Gibson et al., 2015), others reported the absence of a significant effect (Farabolini et al., 2021). Moreover, length seems to differently impact NWR across languages (e.g., 2- to 5-syllable Spanish-like stimuli showed similar complexity as 1- to 4-syllable English-like stimuli in Spanish-English bilinguals; Irizarry-Pérez, Peña, & Bedore, 2021). Hopefully, future research will help clarify this issue.

Then, our results include data of NWR tasks built on different language-specific phonological constraints. Therefore, stimuli with similar characteristics might still differ greatly across languages. Differences such as articulatory complexity or the number of phonemes required to reproduce stimuli might affect NWR performance and its association with language exposure.

In addition, similarity between the languages the child is exposed to might influence said association. For example, bilinguals exposed to languages that are similar (e.g., French and Spanish) might benefit from the prior exposure received in the native language and "use it" to acquire the major language (i.e., language transfer) more than children exposed to languages with lower similarity (e.g., Mandarin and English).

We also point out that the studies included in our work include in their samples children attending the last year of kindergarten and the first years of primary schools. These children are exposed to different degrees to literacy education programs and they have reached different levels of literacy proficiency. Literacy can have an impact on the ability to repeat a sequence of phonemes and, thus, on NWR performance. Mixed evidence has been found on the association between NWR and both literacy and language exposure (Cristia et al., 2020). The effect of literacy on NWR performance should be considered.

A further limitation concerns the statistical nature of subgroup analysis and is the fact that it is a bivariate analysis. Subgroup analysis investigates the effect of a target variable on the studied association without controlling for other variables (e.g., test of interactions among variables to explain variance in NWR performance; Summers et al., 2010). Thus we cannot ensure that results from our subgroup analysis were not influenced by other variables.

Finally, we underline that all included studies were conducted in Western, educated, industrialized, rich, and democratic

(WEIRD) countries. It would be interesting to collect data in non-WEIRD countries to analyze the weight of social and cultural differences in developmental pathways (Muthukrishna, Bell, Henrich, Curtin, Gedranovich, McInerney, & Thue, 2020). A recent work revealed that, in a community of Amazonian villages where infants are rarely spoken to, monolingual kids showed lower NWR scores if compared to data from monolingual children from WEIRD contexts (Cristia et al., 2020).

Conclusions

As conveyed throughout this work, heterogeneity is the keyword to describe differences both between and within bilingual populations, considering each individual's linguistic background, the languages spoken, and the geographical areas. 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, especially as measured by cumulative and current exposure. Further studies should focus on this association in bilingual children with atypical language development, as well as on NWR lists developed with non language-like or cross-linguistic stimuli. Our findings encourage the use of NWR tasks on bilingual children, but researchers and clinicians should be aware that language exposure plays a core role in the NWR performance of this population. Given this task's potential for bilingual language assessment, we hope this work will contribute to a better and deeper understanding of the cognitive and linguistic mechanisms involved in it.

Acknowledgements. The first author contributed to building the research design: he has carried out the literature review, data collection, data extraction and coding, as well as the quantitative meta-analysis; he has also contributed to data interpretation and the writing of the manuscript. The second author has carried out the literature review, data collection and data extraction, and she has contributed to the writing. The third author contributed to the writing of the manuscript and to the evaluation of its coherence and consistency, with an additional role of senior researcher. The last author contributed to the methodological research design, quantitative meta-analysis and data interpretation with an additional role of senior researcher.

Competing interests. Competing interests: The author(s) declare none.

Data availability statement. We present reproducible data [Table S2] and reproducible data code [Appendix S4].

Supplementary Material. For supplementary material accompanying this paper, visit http://dx.doi.org/10.1017/S1366728922000906

Appendix S1. *Prisma flow-diagram.*
Appendix S2. *Inter-rater reliability on literature screening.*
Appendix S3. *Inter-rater data extraction.*
Appendix S4. *Reproducible data for meta-analysis.*
Appendix S5. *Forrest and funnel plots from subgroup analyses.*
Appendix S6. *Forrest and funnel plots from sub-group analysis for risk of bias assessment.*
Table S1. *Literature screening.*
Table S2. *Data extraction.*
Table S3. *Indication for risk of bias.*

References

Altman C, Feldman ZB, Yitzhaki D, Lotem SA and 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. https://doi.org/10.1080/01434632.2013.852561

Anderson NJ, Graham SA, Prime H, Jenkins JM and Madigan S (2021) Linking Quality and Quantity of Parental Linguistic Input to Child Language Skills: A Meta-Analysis. *Child Development* **92**(2), 484–501. https://doi.org/10.1111/cdev.13508

Antonijevic S, Lyons R, Malley M, Meir N, Haman E, Banasik N, Carroll C, McMenamin R, Rodden M and Fitzmaurice Y (2019) Language assessment of monolingual and multilingual children using non-word and sentence repetition tasks. *Clinical Linguistics, & Phonetics* **34**, 1–19. https://doi.org/10.1080/02699206.2019.1637458

Archibald LMD and Gathercole SE (2006) Nonword Repetition: A Comparison of Tests. *Journal of Speech, Language, and Hearing Research* **49**(5), 970–983. https://doi.org/10.1044/1092-4388(2006/070)

Armon-Lotem S, de Jong J and Meir N (2015) *Assessing Multilingual Children: Disentangling Bilingualism from Language Impairment.* Multilingual Matters.

Armon-Lotem S and 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), 715–731. https://doi.org/10.1111/1460-6984.12242

Baddeley A (1986) Working Memory. *Oxford Psychology Series* **11**.

Barbosa PG, Jiang Z and Nicoladis (2017) The role of working and short-term memory in predicting receptive vocabulary in monolingual and sequential bilingual children. *International Journal of Bilingual Education and Bilingualism.* https://doi.org/10.1080/13670050.2017.1314445

Boerma T, Chiat S, Leseman P, Timmermeister M, Wijnen F and Blom E (2015) A Quasi-Universal Nonword Repetition Task as a Diagnostic Tool for Bilingual Children Learning Dutch as a Second Language. *Journal of Speech, Language, and Hearing Research* **58**(6), 1747–1760. https://doi.org/10.1044/2015_jslhr-l-15-0058

Bonifacci P, Barbieri M, Tomassini M and Roch M (2018) In few words: Linguistic gap but adequate narrative structure in preschool bilingual children. *Journal of Child Language* **45**(1), 120–147. https://doi.org/10.1017/S0305000917000149

Borenstein M, Hedges LV, Higgins JPT and Rothstein H (2009) *Introduction to Meta-Analysis | Wiley.* Wiley. https://www.wiley.com/en-us/Introduction+to+Meta+Analysis%2C+2nd+Edition-p-9781119558392

Bortolini U, Arfé B, Caselli CM, Degasperi L, Deevy P and Leonard LB (2006) Clinical markers for specific language impairment in Italian: The contribution of clitics and non-word repetition. *International Journal of Language, & Communication Disorders* **41**(6), 695–712. https://doi.org/10.1080/13682820600570831

Boutron I, Page MJ, Higgins JP, Altman DG, Lundh A and Hróbjartsson A, on behalf of the Cochrane Bias Methods Group (2019) Considering bias and conflicts of interest among the included studies. In *Cochrane Handbook for Systematic Reviews of Interventions* (pp. 177–204). John Wiley and Sons, Ltd. https://doi.org/10.1002/9781119536604.ch7

Brandeker M and 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

Buac M, Gross M and Kaushanskaya M (2016) Predictors of processing-based task performance in bilingual and monolingual children. *Journal of Communication Disorders* **62**, 12–29. https://doi.org/10.1016/j.jcomdis.2016.04.001

Carroll SE (2017) 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 (2015) Non-word repetition. In *Methods for assessing multilingual children: Disentangling bilingualism from language impairment* (pp. 125–150). Multilingual Matters.

Chiat S and 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 JA and Evans JL (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://doi.org/10.1080/13682820601116485

Conti-Ramsden G, Botting N and Faragher B (2001) Psycholinguistic Markers for Specific Language Impairment (SLI). *Journal of Child Psychology and Psychiatry* **42**(6), 741–748. https://doi.org/10.1111/1469-7610.00770

Core C, Chaturvedy S and 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*, 179–185.

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

de Almeida L, Ferré S, Morin E, Prévost P, dos Santos C, Tuller L, Zebib R and Barthez MA (2017) Identification of bilingual children with Specific Language Impairment in France. *Linguistic Approaches to Bilingualism* **7** (3–4), 331–358. https://doi.org/10.1075/lab.15019.alm

Dispaldro M, Leonard LB and Deevy P (2013) Real-Word and Nonword Repetition in Italian-Speaking Children with Specific Language Impairment: A Study of Diagnostic Accuracy. *Journal of Speech, Language, and Hearing Research* **56**(1), 323–336. https://doi.org/10.1044%2F1092-4388(2012%2F11-0304)

Duncan TS and 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

Engel de Abreu PMJ (2011) Working memory in multilingual children: Is there a bilingual effect? *Memory* **19**(5), 529–537. https://doi.org/10.1080/09658211.2011.590504

Engel de Abreu PMJ, Baldassi M, Puglisi ML and Befi-Lopes DM (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)

Farabolini G, Rinaldi P, Caselli MC and Cristia A (2021) Non-word repetition in bilingual children: The role of language exposure, vocabulary scores and environmental factors. *Speech, Language and Hearing* **0**(0), 1–16. https://doi.org/10.1080/2050571X.2021.1879609

Fisher EL (2017) A Systematic Review and Meta-Analysis of Predictors of Expressive-Language Outcomes Among Late Talkers. *Journal of Speech, Language, and Hearing Research* **60**(10), 2935–2948. https://doi.org/10.1044/2017_JSLHR-L-16-0310

Gagarina NV, Klop D, Kunnari S, Tantele K, Välimaa T, Balčiūnienė I, Bohnacker U and Walters J (2012) MAIN: Multilingual assessment instrument for narratives. *ZAS Papers in Linguistics* **56**, 155–155. https://doi.org/10.21248/zaspil.56.2019.414

Gathercole SE and Masoura EV (2005) Contrasting contributions of phonological short-term memory and long-term knowledge to vocabulary learning in a foreign language. *Memory* **13**(3–4), 422–429. https://doi.org/10.1080/09658210344000323

Gatt D and 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

Gibson TA, Summers C, Peña ED, Bedore LM, Gillam RB and Bohman TM (2015) The role of phonological structure and experience in bilingual children's nonword repetition performance. *Bilingualism: Language and Cognition* **18**(3), 551–560. https://psycnet.apa.org/doi/10.1017/S1366728914000248

Graf Estes K, Evans JL and Else-Quest NM (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)

Guiberson M and Rodríguez BL (2015) Nonword Repetition in Spanish-Speaking Toddlers with and without Early Language Delays. *Folia Phoniatrica et Logopaedica* **67**(5), 253–258. https://doi.org/10.1159/000442745

Gutiérrez-Clellen VF and 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%2Fj.1540-5826.2009.00300.x

Haman E, Wodniecka Z, Marecka M, Szewczyk J, Białecka-Pikul M, Otwinowska A, Mieszkowska K, Łuniewska M, Kołak J, Miękisz A, Kacprzak A, Banasik N and Foryś-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**. https://doi.org/10.3389/fpsyg.2017.01444

Higgins JPT, Thompson SG, Deeks JJ, and Altman DG (2003) Measuring inconsistency in meta-analyses. *BMJ (Clinical Research Ed.)* **327**(7414), 557–560. https://doi.org/10.1136/bmj.327.7414.557

Hoff E (2020) Lessons from the study of input effects on bilingual development. *International Journal of Bilingualism* **24**(1), 82–88. https://doi.org/10.1177/1367006918768370

Huls S (2017) *Bilingual Children's Performance on Three Nonword Repetition Tasks: The Role of Language Experience and Ability*. [PhD dissertation, University of Kansas].

Irizarry-Pérez CD, Peña ED and Bedore LM (2021) Phonological predictors of nonword repetition performance in bilingual children. *Journal of Communication Disorders* **94**, 106156. https://doi.org/10.1016/j.jcomdis.2021.106156

Kehoe M, Poulin-Dubois D and Friend M (2021) Within- and Cross-Language Relations Between Phonological Memory, Vocabulary, and Grammar in Bilingual Children. *Journal of Speech, Language, and Hearing Research* **64**(12), 4918-4948. https://doi.org/10.1044/2021_JSLHR-21-00176

Kołak J (2020) *Language environment and vocabulary development of the Polish-English bilingual children in the United Kingdom and Ireland.* [PhD dissertation, Univerisity of Warsaw].

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

Liberati A, Altman DG, Tetzlaff J, Mulrow C, Gøtzsche PC, Ioannidis JPA, Clarke M, Devereaux PJ, Kleijnen J and Moher D (2009) The PRISMA Statement for Reporting Systematic Reviews and Meta-Analyses of Studies That Evaluate Health Care Interventions: Explanation and Elaboration. *PLOS Medicine* **6**(7), e1000100. https://doi.org/10.1371/journal.pmed.1000100

Li'el N (2017) *Assessing Dual Language Learners with Primary Language Impairments- Developing an assessment approach* [master thesis, Curtin University].

Limacher S (2019) *Speech Assessment in Bilingual Children: Relationship between Perceptual Judgments of Accent/Comprehensibility and Formal Test Measures.* [master thesis, University of Alberta].

Loukatou G, Scaff C, Demuth K, Cristia A and Havron N (2021) Child-directed and overheard input from different speakers in two distinct cultures. *Journal of Child Language* 1–20. https://doi.org/10.1017/S0305000921000623

McHugh ML (2012) Interrater reliability: The kappa statistic. *Biochemia Medica* **22**(3), 276–282. https://hrcak.srce.hr/89395

Meir N, Walters J and Armon-Lotem S (2015) Disentangling SLI and bilingualism using sentence repetition tasks: The impact of L1 and L2 properties. *International Journal of Bilingualism* **20**(4), 421–452. https://doi.org/10.1177/1367006915609240

Muthukrishna M, Bell AV, Henrich J, Curtin CM, Gedranovich A, McInerney J and Thue B (2020) Beyond Western, Educated, Industrial, Rich, and Democratic (WEIRD) Psychology: Measuring and Mapping Scales of Cultural and Psychological Distance. *Psychological Science* **31**(6), 678–701. https://doi.org/10.1177/0956797620916782

Öberg L (2020) *Words and non-words: Vocabulary and phonological working memory in Arabic-Swedish-speaking 4–7-year-olds with and without a diagnosis of Developmental Language Disorder* [Uppsala University].

https://doi.org/10.1017/S1366728922000906 Published online by Cambridge University Press

Ortiz JA (2021) Using Nonword Repetition to Identify Language Impairment in Bilingual Children: A Meta-Analysis of Diagnostic Accuracy. *American Journal of Speech-Language Pathology* 1–21. https://doi.org/10.1044/2021_AJSLP-20-00237

Orwin RG (1994) Evaluating coding decisions. In *The handbook of research synthesis* (pp. 139–162). Russell Sage Foundation.

Parra M, Hoff E and 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), 113–125. https://doi.org/10.1016/j.jecp.2010.07.011

Pérez-Navarro JJ, Molinaro N, and 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

Reid H (2019) *Nonword repetition and sequential memory in children aged 7-12 and adults – differential effects of bilingualism, developmental language disorders, and the role of development* [PhD dissertation, McGill Univerisity]. https://escholarship.mcgill.ca/concern/theses/hx11xk101

RevMan, The Cochrane Collaboration (2020) Review Manager (RevMan) [Computer program]. Version 5.4. The Cochrane Collaboration.

Ronderos J, Castilla-Earls A and Marissa Ramos G (2021) Parental beliefs, language practices and language outcomes in Spanish-English bilingual children. *International Journal of Bilingual Education and Bilingualism* **0** (0), 1–22. https://doi.org/10.1080/13670050.2021.1935439

Sahlen B, Reuterskiold-Wagner C, Nettelbladt U and Radeborg K (1999) Non-Word Repetition In Children With Language Impairment-Pitfalls And Possibilities. *International Journal of Language, & Communication Disorders* **34**(3), 337–352. https://doi.org/10.1080/136828299247441

Santos C dos and Ferré S (2016) A Nonword Repetition Task to Assess Bilingual Children's Phonology. *Language Acquisition* **0**(0), 1–14. https://doi.org/10.1080/10489223.2016.1243692

Schraeyen K, Elst WVD, Geudens A, Ghesquière P and Sandra D (2018) Beyond global differences between monolingual and bilingual children on the nonword repetition task: Retention skills for phonemes' identity and serial order. *Bilingualism: Language and Cognition* **21**(2), 403–418. https://doi.org/10.1017/S1366728917000244

Schwob S, Eddé L, Jacquin L, Leboulanger M, Picard M, Oliveira PR and Skoruppa K (2021) Using Nonword Repetition to Identify Developmental Language Disorder in Monolingual and Bilingual Children: A Systematic Review and Meta-Analysis. *Journal of Speech, Language, and Hearing Research* **64**(9), 3578–3593. https://doi.org/10.1044/2021_JSLHR-20-00552

Sharp KM and Gathercole VCM (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/10.1177/0265659012465208

Summers C, Bohman TM, Gillam RB, Peña ED and Bedore LM (2010) Bilingual performance on nonword repetition in Spanish and English. *International Journal of Language, & Communication Disorders* **45**(4), 480–493. https://doi.org/10.3109/13682820903198058

Szewczyk JM, Marecka M, Chiat S and Wodniecka Z (2018) Nonword repetition depends on the frequency of sublexical representations at different grain sizes: Evidence from a multi-factorial analysis. *Cognition* **179**, 23–36. https://doi.org/10.1016/j.cognition.2018.06.002

Talli I and Stavrakaki S (2020) Short-term memory, working memory and linguistic abilities in bilingual children with Developmental Language Disorder. *First Language* **40**(4), 437–460. https://doi.org/10.1177/0142723719886954

Thordardottir E (2008) Language-specific effects of task demands on the manifestation of specific language impairment: A comparison of English and Icelandic. *Journal of Speech, Language, and Hearing Research* **51**(4), 922–937. https://doi.org/10.1044/1092-4388(2008/068)

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), 400–422. https://doi.org/10.1080/13670050.2017.1358695

Thordardottir E and 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 and Juliusdottir AG (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), 411–435. https://doi.org/10.1080/13670050.2012.693062

Tuller L, Hamann C, Chilla S, Ferré S, Morin E, Prevost P, Santos C dos, Ibrahim LA and 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

Vaahtoranta E, Suggate S, Lenhart J and Lenhard W (2021) Language exposure and phonological short-term memory as predictors of majority language vocabulary and phonological awareness in dual language learning. *Bilingualism: Language and Cognition* **24**(2), 319–332. https://doi.org/10.1017/S1366728920000541