Accepted for publication in *Journal of Speech, Language, and Hearing Research*
http://jslhr.pubs.asha.org/

PubMed Central may release the manuscript for public access no sooner than 6 months after final online publication in the journal

A framework for crosslinguistic nonword repetition tests: Effects of bilingualism and socioeconomic status on children's performance

Running head: Crosslinguistic nonword repetition tests

Shula Chiat

Language & Communication Science

City University London

Email: shula.chiat.1@city.ac.uk

Kamila Polišenská

School of Psychological Sciences

The University of Manchester

Email: kamila.polisenska@manchester.ac.uk

Corresponding address:

Shula Chiat

Language & Communication Science

City University London

Northampton Square

London EC1V 0HB, UK

Email: shula.chiat.1@city.ac.uk

Abstract

Purpose: As a recognised indicator of language impairment, nonword repetition has unique potential for distinguishing language impairment from difficulties due to limited experience and knowledge of a language. This study focused on a new Crosslinguistic Nonword Repetition framework (CL-NWR) comprising three tests that vary the phonological characteristics of nonwords in the quest for an assessment that minimises effects of language experience and knowledge, and thereby maximises potential for assessing children with diverse linguistic experience.

Method: The English version of the CL-NWR was administered, with a test of receptive vocabulary, to 4-7-year-old typically developing monolingual and bilingual children (n=21 per group) from mid-high and low socioeconomic (SES) neighbourhoods.

Results: Receptive vocabulary was affected by both bilingualism and neighbourhood SES. In contrast, no effects of bilingualism or neighbourhood SES were found on two of our nonword repetition tests, while the most language-specific test yielded a borderline effect of neighbourhood SES but no effect of bilingualism.

Conclusions: Findings support the potential of the CL-NWR tests for assessing children regardless of lingual/socioeconomic background. They also highlight the importance of considering the characteristics of nonword targets, and investigating the compound influence of bilingualism and SES on different language assessments.

To the extent that nonword repetition is effective in identifying language impairment in monolingual children, it holds particular potential for assessment of bilingual children and others with limited exposure to the target language. Unlike tests of receptive and expressive language, nonword repetition does not require knowledge of lexical semantics and morphosyntax. As a task indicative of language impairment across languages (Chiat, 2015; Gathercole, 2006; Graf-Estes, Evans, & Else-Quest, 2007), and one that is less reliant on language-specific knowledge than other tests, nonword repetition might help to distinguish children with limited language knowledge due to limited experience of the target language from those with language impairment.

However, nonword repetition is not entirely independent of language experience. There is now extensive evidence that children's ability to repeat nonwords is significantly affected by the phonological proximity of the nonwords to real words in the language: children are better able to repeat nonwords that are more wordlike, contain real morphemes of the language, have higher phonotactic probability and/or fall into dense lexical neighbourhoods (Jones, Tamburelli, Watson, Gobet, & Pine, 2010; Leclercq, Maillart, & Majerus, 2013; Messer, Leseman, Boom, & Mayo, 2010; Metsala & Chisholm, 2010). These effects indicate that experience and knowledge of lexical phonology contribute to nonword repetition. Accordingly, nonword repetition performance is generally found to relate to vocabulary knowledge in monolingual children (Gathercole, 2006), so even on nonword repetition tasks, performance may be affected by limited language exposure and/or limited vocabulary.

In line with this evidence, a number of studies of nonword repetition have found significant differences between monolingual and bilingual groups (Cockroft, 2016; Engel de Abreu, 2011; Engel de Abreu, Baldassi, Puglisi, & Befi-Lopes, 2013; Messer et al., 2010; Kohnert, Windsor, & Yim, 2006; Windsor, Kohnert, Lobitz, & Pham, 2010). Engel de Abreu and colleagues (2011, 2013) found that group differences disappeared once vocabulary was controlled, supporting the contribution of language experience and/or knowledge; a recent study by Sorenson Duncan and Paradis (2016) found that length of exposure and characteristics of home language, as well as vocabulary knowledge, affected nonword repetition performance in groups of sequential bilingual children exposed to South Asian or Chinese languages. However, other studies have found no difference between monolingual and bilingual groups (Lee & Gorman, 2012; Lee, Kim, & Yim, 2013; Thordardottir & Juliusdottir, 2013). The few studies investigating socioeconomic (SES) or sociocultural effects have produced more consistent results. Most have found low SES or minority groups performing in line with more advantaged peers on nonword repetition, whilst replicating well-established differences on vocabulary and other measures of language (Balladeres, Marshall, & Griffiths, in press; Campbell, Dollaghan, Needleman, & Janosky, 1997; Engel, Santos, & Gathercole, 2008; Law, McBean, & Rush, 2011). On the other hand, the standardisation of the Preschool Repetition Test (Seeff-Gabriel, Chiat, & Roy, 2008), which includes real words as well as nonwords, revealed significant effects of SES as measured by parental education. These effects arose from the lowest SES group, whose parents had *no* educational qualifications. This finding of differential performance in the lowest SES group was replicated in a large-scale study comparing groups of 3½-5-year-olds living in low SES vs mid-high SES neighbourhoods (Roy & Chiat, 2013), both on the Preschool Repetition Test as a whole, and on the separate sets of real words and nonwords (personal communication).

To our knowledge, the combined effects of language background and socioeconomic status on nonword repetition have not been investigated. However, two recent studies have investigated the effects of these factors on other measures of language. Calvo and Bialystok (2014) assessed receptive vocabulary in four groups of 6-7-year-old children sharing their school and neighbourhood environments, but distinguished by SES (working class vs middle class) and home language status (English monolingual vs bilingual from widely mixed language and cultural backgrounds). This study found significant and independent effects of SES and home language status, with no interaction between these. Gathercole, Kennedy and Thomas (in press) analysed the contribution of home language and SES to performance on Welsh and English receptive vocabulary and grammar across seven age bands ranging from 3 years to older adults in four language groups: English monolingual and Welsh-English bilingual with English-only, Welsh-only or both English and Welsh spoken at home. Again, both language and

SES were found to influence performance, but their contribution varied across age, with some indication that home language was more influential at younger ages and SES at later ages. In discussing their findings, both Calvo and Bialystok (2014) and Gathercole and colleagues (in press) pointed out the possible confounding of socioeconomic and lingual status in studies of bilingual children, and highlighted the importance of investigating the contribution of these combined factors to children's development and performance.

In the case of nonword repetition, effects of lingual and socioeconomic status have varied between studies. These differences in findings could arise from differences between participant groups: their age (varying from 2-6 to 11-14 across studies cited); in the bilingual studies, the particular languages involved in monolingual and bilingual samples, the type and amount of children's exposure to each, and the cultural and socioeconomic status of their languages and communities; and in the SES studies, the nature and extent of the SES differences between groups. However, different findings could also arise from differences between the nonword repetition tests used, and more specifically, the way that items in these tests are constructed, which may in turn influence the effects of participant factors.

If nonwords are made up of syllables that are not real morphemes and have low phonotactic probability in the target language (for example, /dɔɪf/, /teɪvɔɪtʃɑɪg/, /nɑɪtʃɔɪtɑʊvub/, in the Nonword Repetition Test (NRT), Dollaghan & Campbell, 1998), the opportunity to benefit from experience of the language is minimised, and children with a larger vocabulary should only be advantaged to the extent that their vocabulary includes more items of low phonotactic probability. But if nonwords contain real morphemes, particularly morphemes that occur in higher-level vocabulary (for example, *trumpetine*, *stoppagrattic*, *fenneriser*, *versatrationist*, in the Children's Test of Nonword Repetition (CNRep), Gathercole and Baddeley, 1996, and the Portuguese and Luxembourgish tests used by Engel de Abreu and colleagues, 2011, 2013), they will be more familiar to children with more extensive vocabularies and may be easier for them to repeat. The benefits of phonological familiarity were evident in a within-subject comparison of performance (Archibald & Gathercole, 2006) on the CNRep and NRT: percentage consonants correct was significantly higher on the English-like CNRep, even though items in this test go up to five syllables and contain clusters, in contrast to the NRT items which stop at four syllables and contain no clusters.

Test construction has received little attention in the bilingual studies reported above, and information about the phonological characteristics of nonword targets is not always available. Nonetheless, some differences in test construction are evident, and these may affect levels of nonword repetition performance across languages both within and between children. For example, in a study that administered English and Spanish nonword repetition tests to Spanish-English bilingual and English monolingual children, Windsor and colleagues (2010) pointed out that ‘Although the two tasks are parallel, they are not designed to be directly equivalent in item difficulty’ (p. 303). The finding that the Spanish-English bilingual children performed significantly better on the Spanish test relative to their English monolingual peers indicates that these children benefitted from their familiarity with Spanish phonology. Interestingly, though, scores for the monolingual English children on the English and Spanish tests (Table 2, p.304) are very similar, even though the Spanish test contained Spanish realisations of consonants and vowels which were unfamiliar to them, and included longer (five-syllable) items. The English test differed from the Spanish test in two ways that might have made it more challenging: first, the nonwords contained word-final consonants where the Spanish nonwords contained only open syllables, and second, all syllables in the English items contained tense vowels resulting in a prosody which is alien to English words, where the Spanish items followed the Spanish pattern of stress on the penultimate syllable rendering them prosodically more similar to real English words. Either of these phonological characteristics may have made the English test more challenging than the Spanish test for both groups of children, counteracting the benefits of familiar consonant and vowel content for the monolingual group.

Since the characteristics of nonwords are known to influence performance within and between languages, and these can be systematically manipulated, it is important to find out what sort of nonwords maximise the potential of nonword repetition as a clinical assessment for children with heterogeneous language backgrounds and experience of the target language.

The optimal nonword repetition test would be one which shows minimal difference in performance across diverse language experience, and minimal overlap between typically developing children and children with language impairment regardless of language experience. Such a test would be particularly valuable where clinicians working in multilingual communities may have little or no knowledge of children's home language and limited information about their exposure to the majority language.

For these reasons, the issue of test construction was taken up in a European-wide research project addressing the challenges that multilingualism poses for the diagnosis and treatment of language-impaired bilingual children (COST Action IS0804: see http://www.bi-sli.org/). Discussion between colleagues from a wide range of countries and language backgrounds led to the creation of a framework for nonword repetition tests that vary in the proximity of nonwords to the target language.

The Crosslinguistic Nonword Repetition (CL-NWR) framework

The CL-NWR framework (Chiat, 2015) includes three tests:

The crosslinguistic test (CLT) is designed to be maximally compatible with different languages. It comprises 16 nonwords equally divided between lengths 2, 3, 4 and 5 syllables. All syllables are of simple CV structure, made up from a range of consonants (/p, b, t, d, k, g, s, z, l, m, n/), and vowels (/a, i, u/) that were common to all the languages represented in the COST nonword repetition group (see above) and are among the most common phonemes in world languages (Lindblom, 1986; Maddieson, 1984). The framework offers alternatives for each of the 16 items in case a particular segmental target is not present in the target language (for example, /g/ is rare in the Dutch phoneme inventory, Boerma et al., 2015), or one particular item is a real word in the target language. The selected items are produced with even length and pitch apart from the final syllable which is assigned greater length and falling pitch to indicate the end of an utterance. However, consonants and vowels are produced in accordance with the target language, resulting in unavoidable phonetic differences between languages. Examples from the British-English version are /ˈluˌmi/, /ˈmɑˈliˌtu/, /ˈziˈpɑˈliˌdɑ/, /ˈduˈliˈgɑˈsuˌmu/.

While this test avoids segmental targets and phonotactic structures that vary substantially between languages, as we have argued (Chiat, 2015), no test can be entirely neutral between languages. Apart from phonetic differences, the even prosody, CV syllable structure, range of C and V segments, and CV sequences will be more characteristic of some languages than others, and items may therefore be more word-like and easier to repeat in some languages than others. Furthermore, although the test items are nonwords in the target language, they may be real words in a language known to the child, and given the limited range of consonants and vowels, they may contain syllables that are real words or morphemes in the target language. This is evident in the British-English version in which the majority of component syllables (for example /du, mi, mu, si, su, tu/) constitute monosyllabic words.

The prosodically-specific test (PST) comprises the same 16 items as the crosslinguistic test, but in this case, the items are produced with the prosody characteristic of real words of the same length in the target language. For example, primary stress is placed on the syllable that would typically carry that stress in a polysyllabic word. Examples from the British-English version are /ˈlumi/, /ˈmɑlɪˌtu/, /ˈzipəˌlidə/, /ˌdulɪgæˈsumə/.

The language-specific test (LST) contains many more features that are specific to a language, and allows us to manipulate how typical these features are of real words in the language. Items draw on the full inventory of consonants and vowels in the target language; include consonant clusters if allowed in the language; and carry different prosodic patterns where these vary in the language. Test items are divided between high and low phonotactic probability sequences. The British-English version comprises 24 items equally divided between two, three and four syllables, and made up of a wide range of English vowels and consonants. Based on two measures of phonotactic probability (transitional probability and ngram frequency, both derived from the corpus biSubtlex-US, Brysbaert & New, 2009), items at each length were equally divided between high probability (transitional probability: *M* = 11.44, *SD* = 3.19; ngram frequency: *M* = 2.85, *SD* = .78) and low probability (transitional probability: *M* = 6.96 , *SD* = 2.57; ngram frequency *M* = 2.08, *SD* = .52). The difference between high and low probability sets was significant (transitional probability: $t(22)= 3.71$, $p = .001$; ngram frequency:

$t(22) = 2.85$, $p = .009$). We sought to exclude syllables that are real words in English, but note four syllables that are potentially familiar phonological forms (/tɒskəlimə, zʊməlɑ, flɑnəmuzə, lɜsnɒk/), and three that may serve as derivational morphemes (/ˈsænəri, stɒfəli, rɪˈvaɪk/). For each combination of length (2-4 syllables) and phonotactic probability (high/low), there were four items. These were further differentiated by prosody and complexity: three had typical English prosody with no cluster, an initial cluster, or medial cluster (e.g. /ˈrɛfəp, ˈfræʃək, ˈlɜsnɒk/); the remaining item had atypical prosody with no cluster (e.g. /nəˈlɔʃ/).

The three tests in the CL-NWR allow us to investigate which is optimal for assessing children from diverse language backgrounds, whilst at the same time providing a unified nonword repetition assessment that is informative about linguistic factors.

Evaluation of the CL-NWR framework

A recent study (Boerma et al., 2015) compared the performance of Dutch monolingual and mixed bilingual groups, with and without language impairment, on the CLT and an existing Dutch-specific nonword repetition test. The CLT yielded no differences between monolingual and bilingual groups, clear differences between typical and language-impaired groups, and reached good levels of diagnostic accuracy in the bilingual group (sensitivity of 83% and specificity of 93%). These outcomes are in line with the aims of the crosslinguistic assessment. The language-specific test, on the other hand, disadvantaged the typically developing bilingual group and accuracy of diagnosis was reduced in this group for the Dutch-specific test (sensitivity of 63% and specificity of 93%). In this study, lingual status was to some extent confounded with SES, since SES was significantly lower in the bilingual than the monolingual typically developing group. It is therefore possible that SES contributed to the observed group differences on the Dutch-specific test. However, the authors' observation that SES was not correlated with nonword repetition performance in any group suggests that SES was not a key factor.

The present paper reports a small-scale study focusing on the British-English version of the CL-NWR framework (Chiat et al., 2012), including the CLT, PST and LST. Our primary aim was to compare the effects of lingual status (monolingual vs bilingual) and neighbourhood SES (mid-high vs low) on the performance of typically developing children on the three nonword repetition tests. We also compared their performance on a test of receptive vocabulary, since vocabulary knowledge necessarily reflects language experience as well as ability. Based on previous evidence, we expected that (i) both lingual status and neighbourhood SES would have significant and independent effects on vocabulary scores; (ii) neither lingual status nor neighbourhood SES would affect the CLT; (iii) lingual status and neighbourhood SES might affect the LST, with bilingual children living in a low SES neighbourhood being most vulnerable. The presence of typical lexical prosody has been found to facilitate nonword repetition performance in monolingual English and Swedish children (Archibald & Gathercole, 2007; Roy & Chiat, 2004; Sahlén, Reuterskiöld-Wagner, Nettelbladt, & Radeborg, 1999), but this has not been investigated in bilingual or low SES groups, so there were no grounds for predicting whether performance on the PST would be affected by lingual or socioeconomic status.

Our secondary aim was to confirm the sensitivity of our newly developed tasks to two factors found to have robust effects across age and across languages: length (Boerma et al., 2015; Polišenská & Kapalková, 2014) and phonotactic probability (Messer et al., 2010).

Methods

Participants

Once ethical approval for this study was granted by the City University School of Health Sciences Ethics Committee, 42 participants were recruited. These included 24 children aged 4-5 years (M = 61.75 months, SD = 8.87) from an inner London neighbourhood of mid-high SES, 12 monolingual and 12 Spanish-English bilingual; and 18 children aged 4-7 years (M = 73.89 months, SD = 8.87) from an outer London neighbourhood of low SES, 9 monolingual and 9 bilingual, the majority Turkish-English. None of the participants had been clinically referred and no concerns had been expressed about their language development by parents or teachers. The monolingual and bilingual groups did not differ in age ($t(40) = -.72$, $p = .474$), but children in the low SES neighbourhood were significantly older than those in the mid/high SES neighbourhood ($t(40) = 4.42$, $p < .001$). Lingual status was based on parent and teacher report.

Since information about age of onset, length and intensity of exposure to each language was not available for all participants, these factors are not considered in this study (see discussion).

Materials

The British Picture Vocabulary Scales III (BPVS; Dunn, Dunn, Sewell, & Styles, 2009) was administered to assess children's English vocabulary. Nonword repetition was assessed using the British-English version of the CL-NWR tests (Chiat, Polišenská, & Szewczyk, 2012). The full set of items in the three tests is available in the Appendix (re-produced from Chiat, 2015), and the PowerPoint presentation of these is available on request from the authors.

All items were recorded by a female speech and language therapy student with a London accent on a Marantz Professional PMD620 digital recorder. The three tests were embedded in a story presented on PowerPoint (Polišenská & Kapalková, 2014). First, the children were shown a necklace made from colourful beads. On the next slide, the beads appeared spread across the screen and the children were told that the necklace had broken but they could fix it by saying magic words. After the child repeated or attempted to repeat a magic word (i.e. nonword), a bead appeared on the necklace. Every bead appeared with an animated effect and stayed on the screen, allowing the children to see their progress. The experimenter controlled when each nonword was played. At the end of the test, the whole necklace appeared. This was repeated for each nonword repetition test.

Procedure

Children were seen individually in a quiet room. The BPVS was administered first. The CL-NWR tests were then introduced to the children and presented on a laptop through children's headphones. The presentation started with two practice items to familiarise children with the repetition task. The three tests were then administered, with order of presentation counterbalanced across children. Administration was controlled by the researcher and each stimulus was only played once. If the child did not produce a response, the researcher encouraged the child to attempt a repetition but recordings were not replayed even if requested. Children were praised regardless of accuracy, and no feedback was given as to whether the child's response was correct or incorrect. Participants' responses were recorded on a Marantz Professional PMD620 digital recorder. The tester scored responses online and later checked these against the recordings. All recorded responses of all 42 children were also scored independently by a second rater blind to the initial scoring and to age and background of the child, and these scores were used in the analyses.

Scoring

Whole-item scoring was chosen as it is clinically more appropriate and has been found to be informative (Roy & Chiat, 2004; see comparison of scoring methods in Boerma et al., 2015). Responses were scored as correct if all and only phonemes in the target nonword were produced in the correct order. Any phoneme substitutions, omissions and additions were scored as incorrect. Some tests do not penalise addition of phonemes, on the grounds that the child has not lost phonological information (Boerma et al., 2015; Dollaghan & Campbell, 1998). However, phoneme addition can also be seen as evidence that the child has not preserved the phonological input precisely, and in line with our protocol, some tests treat phoneme addition as an error (Sorenson Duncan & Paradis, 2016; Wagner, Torgesen, & Rashotte, 1999). Furthermore, in a breakdown of error types, Burke and Coady (2015) found significantly more additions in children with SLI than typically developing children, though these were greatly outnumbered by substitutions in both groups. Non-responses were scored as incorrect.

Inter-rater reliability

Following Hallgren (2012), inter-rater reliability was evaluated using the intra-class correlation coefficient (ICC). Cicchetti (1994) described values between .75 and 1.0 as excellent. All three nonword sets showed excellent agreement (crosslinguistic test: ICC = .82, prosodically-specific test: ICC = .92, language-specific test: ICC = .90). Overall these results confirm that the scoring system is reliable and fit for purpose.

Results

Group comparisons on vocabulary and nonword repetition tests

Table 1 shows the mean raw score and standard deviation for the BPVS and three nonword repetition tests according to group: monolingual vs bilingual, and low vs mid-high SES. Since neighbourhood groups differed in age, standard scores are also provided for the

BPVS and mean scores adjusted for age are provided for the nonword repetition tests (see below).

The data in Table 1 reveal notable differences in vocabulary performance, with higher standard scores in the monolingual than bilingual group, and in the mid-high than low neighbourhood group. Age-adjusted means for the nonword repetition tests do not reveal such differences.

INSERT TABLE 1 ABOUT HERE

In order to investigate group differences, a two-way between-groups Multivariate Analysis of Covariance (MANCOVA), with age covaried, was performed. The dependent variables were raw scores on the BPVS and the crosslinguistic, prosodically-specific, and language-specific nonword repetition tests. The independent variables were *lingual group* (two levels: monolingual, bilingual) and *neighbourhood group* (two levels: low SES, mid/high SES). The effect size of each result was obtained using the partial eta squared proposed by Cohen (1988), where .01–.05 = small effect, .06–.13 = moderate effect, and a value more than .14 = large effect. Statistical significance was indicated by p < .05 (two-tailed). Using Pillai's trace, there was a significant effect of *lingual group*, V = .31, F(4,34) = 3.78, p =.012, η_p^2 = .31, a significant effect of *neighbourhood group*, V = .42, F(4,34) = 6.11, p = .001, η_p^2 = .42 and a non-significant interaction, V = .10, F(4,34) = .97, p = .436, η_p^2 = .10. The covariate *age* was found to be significant, V =.46, F(4,34) = 7.36, p <.001, η_p^2 = .46.

On the BPVS, the MANCOVA revealed a significant effect of *lingual group* (F(1,37) = 13.40, p = .001, η_p^2 = .27), with the monolingual children gaining higher scores than their bilingual peers, and also *neighbourhood group* (F(1,37) = 25.11, p < .001, η_p^2 =.40), with children from the mid-high SES area gaining higher scores than those from the low SES area. In both cases, effect sizes were large. The interaction effect of *lingual group*neighbourhood group* approached significance for the BPVS score (F(1,37) = 3.95, p = .054, η_p^2 = .10), with moderate effect size, suggesting that the bilingual group from the low SES area might be particularly limited in receptive vocabulary. Indeed, looking at the mean raw scores for the BPVS, it is evident that neighbourhood effects exceed lingual status effects. Even though the low SES groups are on average one year older (mean age just over 6 years, compared with just over 5 years in the mid-high SES group), the mean raw score for the low SES monolingual group is almost the same as for both mid-high SES groups (all 76-77), while the mean raw score for the low SES bilingual group is much lower (just over 49). Taking age into account, the mean standard scores reveal a relatively small gap between the mid-high SES monolingual and bilingual groups (means of 105.67 and 98.17 respectively), both falling in the normal range; in contrast, the gap between the low SES groups is large (means of 90.11 and 74.56 respectively), and the bilingual group mean falls well below the normal range. The monolingual group mean is within the low normal range, but well below even the bilingual group from the mid-high SES neighbourhood (means of 90.11 and 98.17 respectively).

In contrast to findings on the BPVS, no significant effects of lingual group were found on any of the nonword repetition tests (crosslinguistic test: $F(1,37) = .17$, $p = .682$, $\eta_p^2 = .00$; prosodically-specific test: $F(1,37) = 1.07$, $p = .307$, $\eta_p^2 = .03$; language-specific test: $F(1,37) = .84$, $p = .366$, $\eta_p^2 = .02$) or *neighbourhood group* (crosslinguistic test: $F(1,37) = .69$, $p = .411$, $\eta_p^2 = .02$; prosodically-specific test: $F(1,37) = .02$, $p = .876$, $\eta_p^2 = .00$; language-specific test: $F(1,37) = 4.02$, $p = .052$, $\eta_p^2 = .10$). However, the *neighbourhood* difference on the language-specific nonword repetition test approached significance, with moderate effect size, suggesting that children from the low SES neighbourhood, who had poorer receptive vocabulary performance (see above), might have been at a disadvantage repeating nonwords that reflect characteristics of real words in the language. This was consistent across low SES monolingual and bilingual language groups, with no significant interaction between *lingual group* and *neighbourhood group* found for any of the nonword repetition tests (CLT: $F(1,37) = .79$, $p = .381$, $\eta_p^2 = .02$; PST: $F(1,37) = .40$, $p = .533$, $\eta_p^2 = .01$; LST: $F(1,37) = .87$, $p = .356$, $\eta_p^2 = .02$). In line with Engel de Abreu's finding (2011) that group differences disappeared once vocabulary was taken into account, when we controlled for BPVS and age in a two-way ANCOVA of performance on the LST, the borderline significant difference between neighbourhood groups dropped to a non-significant level ($F(1,36) = .01$, $p = .929$, $\eta_p^2 = .00$); *lingual status* ($F(1,36) =

.18, p = .670, η_p^2= .01) and the interaction *SES*lingual status* remained non-significant (F(1,36) = .03, p = .854, η_p^2= .00.

The covariate *age* was significantly related to all four dependent variables, all with large effect sizes: CLT (F(1,37) = 10.67, p = .002, η_p^2 = .22), PST (F(1,37) = 10.85, p = .002, η_p^2 = .23), LST (F(1,37) = 20.74, p < .001, η_p^2 = .36), BPVS (F(1,37) = 18.24, p < .001, η_p^2 = .33).

Effects of phonological factors

As the nonword repetition tasks were newly constructed, we checked that performance on these tasks replicated well-established effects of length and phonotactic probability. A 4 × 2 × 2 mixed-design ANOVA was run with *length* as a within-subject factor with four levels (2, 3, 4, and 5 syllables), and *lingual group* (monolingual, bilingual) and *neighbourhood group* (low SES, mid/high SES) as between-subject factors. The dependent variable was percentage of correctly recalled 2-, 3-, 4- and 5-syllable nonwords, collapsed across the three nonword repetition tasks. The ANOVA revealed a significant main effect of *length*, F (3,123) = 119.35, p <.001, η_p^2 = .74. As expected, as length increased, scores decreased. In line with results of the MANCOVA above, effects of *lingual group* and *neighbourhood group* were non-significant, and no significant interactions were observed (all *ps* < .05).

A 2 × 2 × 2 mixed-design ANOVA was used to evaluate the effects of phonotactic probability on the LST with *phonotactic probability* as a within-subjects factor with two levels (low phonotactic probability, high phonotactic probability), and *lingual group* and *neighbourhood group* as between-subjects factors. The dependent variable was percentage of correctly recalled nonwords of low and high phonotactic probability. The ANOVA found a significant effect of *phonotactic probability*, F (1, 38) = 65.45, p < .001, η_p^2 = .63, with more accurate repetition of nonwords that had higher phonotactic probability. Effects of *lingual group, neighbourhood group* and interactions were again non-significant (all *ps* < .05).

Discussion

As expected, our measure of vocabulary, the BPVS, was strongly affected by experience. Performance of the bilingual children was significantly lower than that of their monolingual peers. Performance was also significantly lower in the low SES than the high SES neighbourhood groups, and with a larger effect size. Accordingly, the interaction between lingual and SES status fell just short of significance, suggesting the bilingual children from the low SES neighbourhood were at a particular disadvantage. Strikingly, the majority of bilingual children in the low SES neighbourhood obtained standard scores that would indicate impairment in a monolingual child (M = 74.56, SD = 6.98). These results are in line with previous studies which have reported independent effects of language background and socioeconomic status on vocabulary performance (Calvo & Bialystok, 2014; Gathercole et al., in press). However, Calvo and Bialystok found no interaction between the two factors in their 6-year-old sample, though the reported effect size is larger for lingual group than SES. Gathercole et al., comparing performance from preschool to adulthood, found that the relative contribution of home language and SES status varied between age groups. Since these studies involved different language combinations and different levels of SES, as well as different age ranges, we might expect different profiles of lingual and SES effects. In our study, SES was confounded with home language (Turkish vs Spanish), so home language may have been a factor in the particularly marked difference between the mid-high and low SES bilingual groups (see Limitations, below, for discussion). As Calvo and Bialystok point out, it is possible that bilingualism and SES 'interact and their effect depends on a specific level of the other' (p.278).

In the light of this, our groups' performance on the CL-NWR tests is of considerable interest. In line with expectations, the CLT showed no effects of either lingual status or neighbourhood SES. Likewise, the PST was unaffected by experience, suggesting that language-specific prosody did not advantage those with greater experience and knowledge of English vocabulary. Given previous findings that typical prosodic structure benefits nonword repetition in monolingual children (Archibald & Gathercole, 2007; Roy & Chiat, 2004; Sahlén et al., 1999), the lack of difference between lingual groups is interesting. It is notable that prosodic effects have been investigated in English and Swedish, both stress-timed languages with vowel reduction in unstressed syllables, and it is possible that prosody is less important in syllable-timed languages. Indeed, the idea of including a prosodically-specific test in the CL-NWR framework may reflect a persisting dominance of English research issues and evidence.

Investigation of prosodic effects across stress- and syllable-timed languages would be of theoretical interest, as well as indicating whether this language-specific factor is worth manipulating for either class of language.

Outcomes for the LST were less clear-cut. There was still no difference between the two lingual groups, but the difference between neighbourhood SES groups fell just short of significance. This profile of findings is surprising when previous studies have more often found effects of lingual status than SES, but since the two factors have not been investigated simultaneously, it is possible that they have been confounded in some studies. Likewise, our finding that lingual status did not affect performance on the LST appears to be at odds with Boerma et al. (2015) whose monolingual and bilingual groups differed significantly on a Dutch language-specific test (though not on the crosslinguistic test). As pointed out above, SES differences between the monolingual and bilingual typically developing groups may have been a factor in their performance on the language-specific test, although correlations between SES and nonword repetition performance were not observed within groups suggesting that observed group differences were not due to SES. The borderline neighbourhood effect on the LST in our study does, however, tally with our finding of a larger effect size for neighbourhood SES than lingual status on the BPVS, pointing to a possible impact of vocabulary knowledge on repetition of more language-specific nonwords. Indeed, when we repeated the analysis with vocabulary controlled, the borderline neighbourhood effect disappeared. This echoes Engel de Abreu and colleagues' report (2011, 2013) that effects of lingual status on their language-specific tests disappeared when vocabulary was controlled.

Turning to phonological factors investigated in our nonword repetition tests, the length effects we observed are consistent with the robust effects of length found in all languages studied to date, with implications for the role of memory (Gathercole, 2006; Boerma et al., 2015; Polišenská & Kapalková, 2014). The effects of phonotactic probability we observed in all groups suggest that children from diverse backgrounds benefit from knowledge of lexical phonology in the test language, despite wide discrepancies in their lexical knowledge. However, it is possible that phonotactic probability in the test language overlaps with phonotactic probability in children's home language and that phonotactic effects could then arise from knowledge of the home language, as pointed out by Messer et al. (2010) when they found that monolingual Dutch children with *no* knowledge of Turkish showed phonotactic probability effects on a Turkish nonword repetition test. Since phonotactic probability is derived from frequencies of co-occurrence *within a language*, it was systematically manipulated in our language-specific items, but was not considered in the items of the CLT and PST. Phonotactic frequency of these crosslinguistic items could nonetheless be calculated for a given language, and the effects on children's performance investigated. In the light of the phonotactic effects observed across many studies and languages, we would expect to find phonotactic effects on the CLT and PST, assuming that the items in these tests vary sufficiently in phonotactic probability for the language(s) spoken by the child. In future research, it would be interesting to see how phonotactic probability of items in the CLT and PST compares with phonotactic probability of items in the LST, and how this affects children's performance within as well as across the three tests.

Limitations and further research

In considering our findings, several limitations in our sampling must be taken into account. First, although our study only included children about whom no concerns had been raised, our assessment data are limited and we cannot rule out the possibility that some children had unrecognised and undiagnosed difficulties. However, further assessment data would not necessarily solve this problem, given that performance on language tests is known to be affected by socioeconomic and lingual status, with children from bilingual and/or low SES backgrounds at increased risk of performing in the impaired range – the very problem that our study set out to address.

Second, when comparing combined effects of lingual status and neighbourhood SES, numbers in each group were small (either 9 or 12), limiting power to detect group differences. With larger groups, the *neighbourhood group* effects on LST and the *lingual*neighbourhood* interaction effect on BPVS that fell just short of significance, and achieved moderate levels for partial eta squared, might cross the boundary to significance. In the case of other non-significant findings, effect sizes were negligible, so increased power would be less likely to affect outcomes.

Further limitations, pointed out above, were the confounding of home language (Spanish/Turkish) with SES (mid-high/low) in our bilingual group, and the lack of information about age of onset, length and intensity of their exposure to English. These limitations must be borne in mind in considering our findings on the relative and combined effects of lingual status and SES. First, we cannot disentangle the contribution of home language and SES. Second, it is possible that variations in home language and/or exposure to English might have masked effects of bilingual status on nonword repetition. However, effects of bilingual status were evident on our measure of English receptive vocabulary. Furthermore, as both Calvo and Bialystok (2014) and Gathercole et al. (in press) point out, cultural, geographic, and socioeconomic characteristics of language communities are often correlated in the real world. This lends ecological validity to our study, and underlines the need for assessment tools that minimise the influence of these factors.

Since the CL-NWR was motivated by this real-world scenario, our finding of similar performance across lingual/SES groups, in the face of marked differences in vocabulary, is encouraging. We now need to replicate this study with large, age-matched groups of children from typologically diverse language backgrounds, both low and mid-high SES, and with known and varied exposure to English, including recent immigrants with minimal or no exposure to English. Validation of the tests will also require evidence of test-retest reliability, and comparison with performance on established nonword repetition tests. Comparison with the Dollaghan and Campbell (1998) and Gathercole and Baddeley (1996) tests will be particularly interesting given their very different phonological properties outlined above.

Most crucial, however, is the extent to which the tests differentiate children with language impairment. Encouragingly, Boerma et al. (2015) found that the Dutch version of our crosslinguistic test fully differentiated groups of children with and without language impairment, bilingual as well as monolingual. However, as Boerma et al. discuss, their cut-off for language impairment was low compared with many other studies, and differentiation may be less clear in a wider population that includes children with less marked language deficits.

Even if nonword repetition tests have special potential in bilingual assessment, it should be emphasised that they are not sufficient for identifying children with language impairment, whether monolingual or bilingual. First, nonword repetition tests are not entirely accurate in identifying monolingual children with language impairment (Gathercole, 2006; Elllis Weismer et al., 2000). Second, previous research indicates that clinical discrimination is better for a more language-specific test (Archibald & Gathercole, 2006; Graf Estes et al., 2007) which is less appropriate for children with limited language experience. In any case, the scope and limits of nonword repetition should be recognised. Repeating a nonword does not require the mapping of phonology onto semantics, or the longer-term storage of phonology and the phonology-semantics mapping, all necessary for acquiring words and morphosyntax and therefore possible sources of language deficits (see Chiat, 2001). Nor does nonword repetition require social cognitive skills that are crucial for social communication, often impaired in children diagnosed with language impairment (Chiat & Roy, 2013). So, while further research may build on the promise of the CL-NWR tests presented in this paper, research is also needed on assessments that detect other deficits underlying language impairment in children and may be similarly unaffected by language-specific experience.

Conclusions and clinical implications

Overall, our findings are promising for the use of the CL-NWR tests in the assessment of children from diverse linguistic and sociocultural backgrounds. Whilst acknowledging limitations of our samples, the contrast between the groups' similar profile of scores on the crosslinguistic and prosodically-specific nonword repetition tests, and their widely dispersed scores on receptive vocabulary, suggests that these tests may be unaffected by language experience. Performance on the language-specific test also showed minimal effects of experience compared with the vocabulary test, but did reveal a marginal disadvantage for children from the low SES neighbourhood. Notably, these children had marked shortfalls on the vocabulary test. Based on these findings, as well as those reported by Boerma et al. (2015), the crosslinguistic test appears to have good potential for identifying deficits, while the language-specific test might be indicative of children's experience of and proficiency in the test language as well as their language-processing ability. The different tests might therefore provide complementary information about children's abilities and knowledge. The full set of 56 items was quick and easy to administer. Most children enjoyed the presentation and readily completed the task, making this a clinically realistic tool.

We have highlighted the compound effects of lingual, sociocultural, and socioeconomic factors on children's language performance, all of which may contribute to a child's low performance. We have argued that variations in the construction of nonword repetition tests allow us to control the contribution of these factors. If our findings are replicated in further investigations with typically developing children, and if the tests distinguish children with language impairment, the CL-NWR will offer a valuable new tool for clinical assessment of children in diverse communities.

Acknowledgements

We thank Roshni Patel and Omobolanle Ogunbanke for recording the CL-NWR tests; Roshni Patel; and Adriana Sanchez for data collection, and Amna Arfin-Hyder for inter-rater reliability checks. The development of the nonword repetition tests reported in this paper was supported by COST Action IS0804 “Language Impairment in a Multilingual Society: Linguistic Patterns and the Road to Assessment” (www.bi-sli.org).

References

Archibald, L. M. D., & Gathercole, S. E. (2006). Nonword repetition: A comparison of tests. *Journal of Speech, Language, and Hearing Research, 49,* 970-983.

Archibald, L. M. D., & Gathercole, S. E. (2007). Nonword repetition in specific language impairment: More than a phonological short-term memory deficit. *Psychonomic Bulletin & Review, 14,* 919-924.

Balladeres, J., Marshall, C., & Griffiths, Y. M. (in press). Socio-economic status affects sentence repetition, but not non-word repetition, in Chilean pre-schoolers. *First Language*.

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

Brysbaert, M., & New, B. (2009). Moving beyond Kučera and Francis: A critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for American English. *Behavior Research Methods, 41,* 977-990.

Burke, H.L., & Coady, J.A. (2015) Nonword repetition errors of children with and without specific language impairments (SLI). *International Journal of Language & Communication Disorders, 50,* 337-346.

Calvo, A., & Bialystok, E. (2014). Independent effects of bilingualism and socioeconomic status on language ability and executive functioning. *Cognition*, *130*(3), 278-288.

Campbell, T., Dollaghan, C., Needleman, H., & Janosky, J. (1997). Reducing bias in language assessment: Processing-dependent measures. *Journal of Speech, Language, and Hearing Research, 40*, 519-525.

Chiat, S. (2001). Mapping theories of developmental language impairment: Premises, predictions and evidence. *Language and Cognitive Processes, 16,* 113-42.

Chiat, S. (2015). Nonword repetition. In Armon-Lotem, S., de Jong, J., & Meir, N. (Eds.) *Methods for assessing multilingual children: Disentangling bilingualism from language impairment*. Bristol: Multilingual Matters, pp.125-150.

Chiat, S., & Roy, P. (2013). Early predictors of language and social communication impairments at 9-11 years: A follow-up study of early-referred children**.** *Journal of Speech, Language, and Hearing Research, 56,* 1824-36.

Chiat, S., Polišenská, K., & Szewczyk, J. (2012). *Crosslinguistic Nonword Repetition Tasks: British English version. A part of LITMUS COST IS0804 Battery*. Unpublished material.

Cicchetti, D. V. (1994). Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychological Assessment, 6,* 284-290.

Cockcroft, K. (2016). A comparison between verbal working memory and vocabulary in bilingual and monolingual South African school beginners: implications for bilingual language assessment. *International Journal of Bilingual Education and Bilingualism*, *19*(1), 74-88.

Cohen, J. (1988), *Statistical Power Analysis for the Behavioral Sciences, 2nd Edition*. Hillsdale, N.J.: Lawrence Erlbaum.

Dollaghan, C., & Campbell, T. F. (1998). Nonword repetition and child language impairment. *Journal of Speech, Language, and Hearing Research, 41,* 1136-1146.

Dunn, L. M., Dunn, D. M., Styles, B., & Sewell, J. (2009).*The British Picture Vocabulary Scale III – 3rd Edition*. London: GL Assessment.

Elllis Weismer, S., Tomblin, J.B., Zhang, X., Buckwalter, P., Chynoweth, J.G., & Jones, M. (2000). Nonword repetition performance in school-age children with and without language impairment. *Journal of Speech, Language, & Hearing Research, 43,* 865-878.

Engel de Abreu, P. M. J. (2011). Working memory in multilingual children: Is there a bilingual effect? *Memory, 19,* 529-537.

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

Engel, P. M. J., Santos, F. H., & Gathercole, S. E. (2008). Are working memory measures free of socioeconomic influence? *Journal of Speech, Language and Hearing Research, 51*, 1580-1587.

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

Gathercole, S. E., & Baddeley, A. D. (1996). *The Children's Test of Nonword Repetition*. London: Psychological Corporation.

Gathercole, V. C. M., Kennedy, I., & Thomas, E. M. (in press). Socioeconomic level and bilinguals' performance on language and cognitive measures. *Bilingualism: Language and Cognition.*

Graf Estes, K., Evans, J., & 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, 5,* 177-195.

Hallgren, K. A. (2012). Computing inter-rater reliability for observational data: An overview and tutorial. *Tutorials in Quantitative Methods for Psychology, 8,* 23-34.

Jones, G., Tamburelli, M., Watson, S.E., Gobet, F., & Pine, J.M. (2010). Lexicality and frequency in Specific Language Impairment: Accuracy and error data from two nonword repetition tests. *Journal of Speech, Language, and Hearing Research*, *53,* 1642-1655.

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.

Law, J., McBean, K., & Rush, R. (2011). Communication skills in a population of primary school-aged children raised in an area of pronounced social disadvantage. *International Journal of Language & Communication Disorders*, *46*(6), 657-664.

Leclercq, A-L., Maillart, C., & Majerus, S. (2013). Nonword repetition problems in children with specific language impairment: A deficit in accessing long-term linguistic representations? *Topics in Language Disorders, 33,* 238-254.

Lee, S. A. S., & Gorman, B.K. (2012). Nonword repetition performance and related factors in children representing four linguistic groups. *International Journal of Bilingualism*, *17,* 479-495.

Lee, H. J., Kim, Y.T., & Yim, D. (2013). Non-word repetition performance in Korean-English bilingual children. *International Journal of Speech-Language Pathology*, *15,* 375-382.

Lindblom, B. (1986). Phonetic universals in vowel systems. In J. Ohala & J. Jaeger (Eds.), *Experimental phonology* (pp. 13-44). Orlando, FL: Academic.

Maddieson, I. (1984). *Patterns of sounds. Cambridge studies in speech science and communication*. Cambridge: Cambridge University Press.

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

Metsala, J. L. & Chisholm, G. M. (2010). The influence of lexical status and neighbourhood density on children's nonword repetition. *Applied Psycholinguistics, 31,* 489-506.

Polišenská, K., & Kapalková, S. (2014). Improving child compliance on a computer administered nonword repetition task. *Journal of Speech, Language, and Hearing Research*, *57*, 1060-1068.

Roy, P., & Chiat, S. (2004). A prosodically controlled word and nonword repetition task for 2- to 4-year-olds: Evidence from typically developing children. *Journal of Speech, Language, and Hearing Research, 47,* 223-234.

Roy, P., & Chiat, S. (2013). Teasing apart disadvantage from disorder: The case of poor language. In C.R. Marshall (Ed.), *Current Issues in Developmental Disorders* (pp.125-150). Hove, East Sussex: Psychology Press.

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

Seeff-Gabriel, B. K., Chiat, S., & Roy, P. (2008). *Early repetition battery*. London: Pearson.

Sorenson Duncan, T., & 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,* 39-48.

Thordardottir, E. T., & Juliusdottir, A.G. (2013). Icelandic as a second language: a longitudinal study of language knowledge and processing by school-age children. *International Journal of Bilingual Education and Bilingualism, 16,* 411-435.

Wagner, R., Torgesen, J., & Rashotte, C. (1999). *Comprehensive Test of Phonological Processing*. Austin, TX: Pro-Ed.

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

Table 1. Mean scores (raw and standard/age-adjusted) and standard deviations for BPVS, CLT, PST, and LST according to lingual and neighbourhood group

Neighbourhood group	**Lingual group**	**BPVS**		**CLT Percentage**		**PST Percentage**		**LST Percentage**	
		Raw Mean (SD)	Standard Mean (SD)	Raw Mean (SD)	Age-adjusted Mean (SE)	Raw Mean (SD)	Age-adjusted Mean (SE)	Raw Mean (SD)	Age-adjusted Mean (SE)
Low SES	**Monolingual** (n=9)	76.78 (17.32)	90.11 (11.63)	67.36 (20.20)	59.46 (5.74)	65.97 (20.28)	57.77 (5.90)	55.56 (20.20)	45.13 (5.42)
	Bilingual (n=9)	49.78 (19.85)	74.56 (6.98)	58.33 (17.40)	52.98 (5.45)	65.28 (17.71)	59.72 (5.61)	43.52 (21.96)	36.46 (5.15)
Mid-high SES	**Monolingual** (n=12)	76.33 (16.46)	105.67 (9.33)	52.08 (15.15)	60.01 (5.11)	45.31 (16.67)	53.53 (5.26)	41.67 (17.23)	52.11 (4.84)
	Bilingual (n=12)	76.17 (11.74)	98.17 (9.90)	60.42 (17.54)	62.44 (4.55)	59.90 (17.77)	62 (4.68)	49.65 (14.04)	52.32 (4.30)

Note: BPVS: British Picture Vocabulary Scales; CLT: Crosslinguistic nonword repetition test; PST: Prosodically-specific nonword repetition test; LST: Language-specific nonword repetition test

Appendix: Crosslinguistic Nonword Repetition Framework: British-English version (Chiat, Polišenská & Szewczyk, 2012, re-produced from Chiat, 2015, pp.149-150)

	Crosslinguistic Test	**Prosodically-Specific Test**	**Language-Specific Test**						
			Stress	Complexity	Length	Transitional probability / Ngram frequency (TP / NF)[1]			
Length						High	TP / NF	Low	TP / NF
2 syll	ˈziˌbu	ˈzibə	Typical	No cluster	2 syll	ˈdælən	11 / 4.56	ˈrɛfəp	8.8 / 3.12
	ˈduˌlɑ	ˈdulə			3 syll	ˈsænəri	16.8 / 3.42	ˈzʊməlɑ	4.6 / 2
	ˈnɑˌgi	ˈnɑgi			4 syll	ˌpɒnəˈveɪkə	13.9 / 2.5	ˌkɛfəˈmɔɪpə	10.7 / 1.6
	ˈluˌmi	ˈlumi		Initial cluster	2 syll	ˈspɒdəl	9.6 / 2.96	ˈfræʃək	7.1 / 2.72
3 syll	ˈsiˈpuˌlɑ	ˈsipəˌlɑ			3 syll	ˈstɒfəli	15.9 / 2.79	ˈsmɪʃəˌtɑʊ	9.5 / 2.28
	ˈbɑˈmuˌdi	ˈbɑməˌdi			4 syll	ˌskuməˈkaɪdə	10.4 / 1.94	ˌflɑnəˈmuzə	8.8 / 1.79
	ˈmɑˈliˌtu	ˈmɑlɪˌtu		Medial cluster	2 syll	ˈnɑskət	12 / 3.15	ˈlɜsnɒk	5.6 / 2.44
	ˈluˈmiˌgɑ	ˈlumɪˌgɑ			3 syll	ˈmɑspəˌdɑʊ	8.8 / 2.25	ˈzispəˌgɔɪ	4.8 / 1.35
4 syll	ˈziˈpɑˈliˌdɑ	ˈzipəˌlidə			4 syll	ˌtɒskəˈlimə	8.7 / 2.1	ˌvɒsnəˈrɑʊdə	7.6 / 1.81
	ˈmuˈkiˈtɑˌlɑ	ˈmukɪˌtɑlə	Atypical	No cluster	2 syll	rɪˈvaɪk	12 / 3.63	nəˈlɔʃ	3 / 2.28
	ˈkɑˈsuˈluˌmi	ˌkɑsəˈlumi			3 syll	pəˈzeɪnə	11.7 / 2.88	ləˈvugə	9.4 / 2.05
	ˈliˈdiˈsɑˌku	ˌlidɪˈsɑkə			4 syll	rəˈnusədɑ	5.3 / 1.96	zəˈdɑgənɜ	3.6 / 1.49
5 syll	ˈsiˈpuˈmɑˈkiˌlɑ	ˌsipəmæˈkilə							
	ˈduˈliˈgɑˈsuˌmu	ˌdulɪgæˈsumə							
	ˈmɑˈluˈziˈguˌbɑ	ˌmɑləˈzigəbə							
	ˈliˈtɑˈpiˈmuˌti	ˌlitəˈpiməti							

[1]Transitional probability and Ngram frequency derived from the corpus biSubtlex-US (Brysbaert & New, 2009)