Non-word repetition

1 author:

Shula Chiat
City, University of London

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

Nonword repetition

Shula Chiat

In the quest for assessments that will address the challenges of diagnosing language impairment in bilingual children, nonword repetition appears to have an advantage over other assessment tasks. In nonword repetition tasks, children are asked to repeat items that, by definition, they have not heard before and are not part of their learned vocabulary. From this, we might infer that children cannot benefit from stored knowledge of the target item, and that children will therefore not be disadvantaged if they have had limited exposure to the target language and have limited knowledge of words in the language. A small body of research has found that differences in language experience amongst monolingual children have less effect on nonword repetition tests than on tests of vocabulary and grammar (Campbell et al., 1997; Engel et al., 2008; Roy & Chiat, 2013).

At the same time, extensive research on nonword repetition in English has found that children diagnosed with primary or specific language impairment (SLI) show deficits in nonword repetition, leading to the proposal that nonword repetition may serve as a clinical marker of language impairment in children (Bishop et al., 1996; Conti-Ramsden et al., 2001). Similar findings have emerged from studies in many languages, often typologically distinct, for example, Italian (Casalini et al., 2007; Dispaldro, Leonard & Deevy, 2013), Dutch (de Bree et al., 2007), Spanish (Girbau & Schwartz, 2007), Russian (Kavitskaya et al., 2011), Swedish (Sahlén et al., 1999), Icelandic (Thordardottir, 2008), French (Thordardottir & Brandeker, 2013). The only exception to date is one study of Cantonese which found no difference between groups of children diagnosed with SLI and chronologically age-matched typically developing children (Stokes et al., 2006).

To the extent that nonword repetition is less affected by language knowledge than other language assessments, and is effective in identifying children with language impairment, it offers a unique tool for diagnosing language impairment in children with limited experience and knowledge of the target language. However, these qualifications are important. While nonword repetition, unlike other language assessments, does not draw directly on knowledge of vocabulary and syntax, it is still influenced by language-specific knowledge. The evidence is that children are better able to repeat nonwords that share phonological characteristics of real words in their language. Since such knowledge necessarily relies on language exposure and experience, we might expect bilingual children to vary in their nonword repetition performance depending on their familiarity with lexical phonology in the target language. Findings to date are mixed.

Some studies find no effects of language experience. Thordardottir and Juliusdottir (2013), for example, report that an L2 Icelandic group (aged just under 5 to just over 17 years) gained very high scores on an Icelandic nonword repetition test, in contrast to their very low scores on a measure of language. Armon-Lotem et al. (in preparation) found that 4-6-year-olds with L1 Russian and L2 Hebrew did not differ from their monolingual peers. Likewise, Lee and Gorman (2012) found that three groups of bilingual 7-year-olds, with L1/Korean, L1/Chinese and L1/Spanish, and varying length of exposure to English, achieved overall accuracy on an English nonword repetition test comparable to that of monolingual English-speaking peers. Lee et al. (2013) investigated performance on a Korean nonword repetition test in monolingual 3-5-year-olds living in Korea and bilingual peers with L1 Korean living in the USA, and again found no significant differences between groups. Thordardottir and Brandeker's (2013) study of nonword repetition in monolingual and bilingual 5-year-olds acquiring French and English simultaneously, but with varied levels of exposure to the two languages, revealed very little effect of language experience on the bilingual children's performance. Furthermore, this study found good differentiation between typically developing bilingual children and monolingual children with language impairment.n contrast, other studies of nonword repetition in children's L1 and/or L2 have found that language experience does affect nonword repetition. Sharp and Gathercole (2013) compared the performance of Welsh-English bilingual children with varying levels of exposure to Welsh in the home and found that children's performance, especially on sounds unique to Welsh, was affected by amount of exposure to the language. Likewise, Lee and Gorman (2012) report that sound inventory (both of consonants and vowels) in L1 is linked to children`s ability to repeat nonwords.

Engel de Abreu (2011) found that bilingual children, with a variety of European languages as L1, performed less well than monolingual children on a Luxembourgish test of nonword repetition. Interestingly, though, the group difference disappeared once vocabulary was controlled, suggesting that reduced language experience was responsible for the difference in nonword repetition. Messer et al. (2010) found that Turkish-Dutch 4-year-olds had lower scores than their Dutch monolingual peers on a Dutch nonword repetition test, but higher scores on a Turkish test, reflecting differences in the two groups' language experience. Studies of the Hispanic population in the USA have been most consistent in finding language and/or group differences. Summers et al. (2010) administered Spanish and English nonword repetition tests to typically developing Spanish-English bilingual children aged 4;6-6;5 and found performance to be significantly better in Spanish. However, comparison of performance across languages must be made with caution, since tests in the two languages are not necessarily matched in phonological difficulty (see further discussion below). Comparing bilingual with monolingual performance on the same English test, Kohnert et al. (2006) and Windsor et al. (2010) found that bilingual Spanish-English children (age ranges 7;10-13;11 and 6;0-11;6 respectively) performed significantly below monolingual English children on an English nonword repetition test. The typically developing bilingual groups nevertheless outperformed the groups with language impairment, monolingual as well as bilingual. This finding is encouraging for the diagnostic potential of nonword repetition. However, this potential is greater where bilingual groups are found to perform on a par with monolingual peers, as in Thordardottir and Brandeker's (2013) study.

This heterogeneity in findings for bilingual children is unsurprising, given the multiple differences between studies. These involve linguistic factors in terms of the particular languages studied and the ways in which nonword repetition tests are constructed in those languages. Social and individual factors also vary between studies, with differences in socioeconomic and cultural characteristics of bilingual communities, in status of children's L1 and L2, and in children's age, length and intensity of exposure to L2. Unpicking the contribution of these factors will require many studies. This chapter focuses on just one of these factors, the construction of nonword repetition tests.

Work on nonword repetition within COST Action IS0804 has given us new insights into the types of language-specific knowledge that may contribute to nonword repetition, and the need to consider these when constructing nonword repetition tasks. Combining previous research findings with our consideration of linguistic factors relevant to nonword repetition, we have developed a framework for designing nonword repetition tests that may be applied across typologically diverse languages, with appropriate adaptations, and with careful attention to the possible impact of children's first language and their exposure to the target language if this is not their first language.

In considering the potential of these tests for diagnosing language impairment in bilingual children and children with limited exposure to the target language, a further note of caution must be sounded. By and large, studies of nonword repetition have compared performance in typically developing (TD) children and children already diagnosed with language impairment, and almost always aged at least 4 years if not older. Finding that a nonword repetition test differentiates between these groups, and reveals a deficit in the LI group, is important. But even in studies of already-defined groups, there is usually some overlap. In the few studies of nonword repetition in a population rather than in defined groups, the overlap increases. So, for example, a test that yielded excellent discrimination between already identified TD and SLI groups (Dollaghan & Campbell, 1998) did much less well in a large population study (Ellis Weismer et al., 2000). In the case of bilingual children, this may be further compounded by differences in language experience. Indeed, Gutiérrez-Clellen and Simon-Cereijido (2010) found that diagnostic accuracy in Spanish-English bilingual children increased if performance on both English and Spanish nonword repetition tests was considered. Thus, while children with language impairment are at risk of nonword repetition deficits, nonword repetition performance is not a litmus test for LI, a qualification that must be borne in mind when using nonword repetition as a diagnostic tool in both monolingual and bilingual children.

Having highlighted the unique potential of nonword repetition for cross-linguistic clinical assessment, as well as reasons for caution in its use, this chapter presents the nonword repetition framework we have developed. The chapter starts with a brief review of the nature of nonword repetition tasks and the factors that influence performance, the relevance of these factors across different languages, and the characteristics of nonword repetition tasks that make them more or less effective in differentiating children with and without LI. This provides the background and motivation for the COST IS0804 nonword repetition framework. The aim of the framework, fundamental to the project of developing assessments for bilingual children, is to identify tests that on the one hand maximise the gap between performance of children with and without LI, and on the other hand minimise the gap between typically developing children with more vs less experience of the target language. In order to explore which types of nonwords achieve the best trade-off between these objectives, our framework incorporates three sets of items that vary in proximity to the phonology of the target language:

1. *Quasi-universal items with quasi-neutral prosody:* Phonological content is selected to be compatible with phonologies of diverse languages.
2. *Quasi-universal items with language-specific prosody:* The items are the same as (1), but with language-specific prosody applied.
3. *Language-specific items:* Phonological content is compatible with and representative of lexical phonology in the target language, with systematic manipulation of phonological factors known to influence performance in monolingual children.

Theoretical and empirical influences on nonword repetition test development

Factors influencing nonword repetition performance in typically developing children

The first factor that was observed to influence nonword repetition performance was item length. Effects of length, indicating declines in performance as item length increased, were taken as evidence that nonword repetition was a measure of short-term memory capacity. However, it was not long before the influence of prosodic and segmental structure of items became evident. Furthermore, the influence of these factors depended in part on the characteristics of the target language. These findings led to nonword repetition being considered a test of phonological abilities rather than pure memory. Some would argue that the two are inseparable, in that memory capacity will depend on the strength of processing of the material to be stored in memory. A keynote paper by Gathercole (2006) together with commentaries by many researchers in the field provides a good overview of debates about the nature of nonword repetition and evidence bearing on these. For our purposes of test development, we look to this research for evidence of characteristics of nonwords that influence performance and are informative about children's language abilities.

Length: Length effects have been observed in many languages apart from English, including Cantonese (Stokes et al., 2006), Hebrew and Russian (Armon-Lotem et al., in preparation). To our knowledge they have been found in every language in which length of nonwords has been considered.

Prosodic structure: Few studies have investigated the effects of prosodic structure, but those that have done so have found significant effects on performance in Swedish (Sahlén et al., 1999) and English (Roy & Chiat, 2004; Chiat & Roy, 2007; Williams et al., 2013).These studies show children having particular difficulties with initial unstressed syllables which are atypical in words of their language. The broader effects of prosody are most clearly revealed in a study by Archibald and Gathercole (2007) which compared repetition of the same syllable sequences in two conditions: in one, the sequence of syllables was produced as a list, with even stress on each (e.g. *fow . . .moy . . . chee*); in the other, the syllables were produced with the rhythmic pattern of a single nonword (e.g. *fowmoychee*). The only difference between conditions, then, was in the prosody assigned to the string of syllables, and this was found to facilitate children's repetition.

Segmental complexity: Segmental complexity has emerged as a key factor, with children's performance consistently better for nonwords containing only single consonants compared with nonwords containing clusters (English: Archibald & Gathercole, 2006; Jones, Tamburelli et al., 2010; Hebrew and Russian: Armon-Lotem et al., in preparation).

Phonotactic probability: A number of studies have established that nonword repetition is affected by the relative frequency of phoneme sequences within items, with performance better for items that have higher phonotactic probability (Jones et al., 2010; Munson, 2001; Munson et al., 2005). It should be pointed out that calculation of phonotactic probability is normally derived from bigram frequencies, that is, frequency of each pair of adjacent phonemes making up the nonword. Frequency of co-occurrence can be calculated for longer sequences, for example trigrams. The most comprehensive measure of phonotactic probability, n-gram frequency, takes into account frequency of co-occurrence of successively longer sequences of phonemes in an item, from the smallest (bigrams) to the largest (full sequence of phonemes making up the item). It should also be noted that phonotactic probability does not take into account the position of phonemes within the prosodic and syllabic structure of the word, for example whether two phonemes form an initial cluster of a stressed or unstressed syllable or a sequence of singleton coda and singleton onset crossing a syllable boundary.

The properties of nonwords that have been systematically investigated reflect the lexical typology of languages studied to date (largely Germanic and Romance). Length and segmental complexity are the most consistently represented variables in these language families. Prosody is more variable: some languages such as English, Swedish, Greek and Russian have variable word stress and therefore more vs less typical stress patterns; others have fixed stress, for example, French, Finnish, Hebrew and Polish. Furthermore, some languages reduce vowels in unstressed syllables, as in English, Swedish and German, while others preserve vowel quality, as in Greek, Italian, and Spanish. It is striking that prosodic effects on nonword repetition have been investigated in English and Swedish, but not to our knowledge in Romance languages. It follows from these observations that we may have overlooked potentially influential factors that do not arise in the language families studied to date. A striking example is *tone*, which is a critical feature of lexical phonology in Chinese languages. In Stokes et al.'s (2006) study of Cantonese, nonwords at each length were assigned a single tonal pattern that was characteristic of real words of the same length, hence tone was not manipulated in this study.

All the factors we have considered above are objective properties of lexical phonology. One other characteristic of nonwords that has been investigated is *wordlikeness*. This is a measure of native speakers' judgement about the extent to which a nonword resembles a real word. It is therefore a subjective rating rather than an inherent property of the nonword. We might assume that the objective properties of nonwords we have considered will contribute to these subjective judgements of wordlikeness: that items with more typical length and prosody, and higher frequency phoneme sequences, will be judged more wordlike, and these factors are indeed found to correlate with wordlikeness judgements (Polišenská et al., in preparation). However, other factors may contribute, for example whether items contain syllables or syllable sequences that are real morphemes, which may be whole words or affixes (see Casalini et al., 2007, and below). Comparison of children's performance on more vs less wordlike items has revealed significant effects (Gathercole, 1995). Casalini et al. report better performance on items containing Italian roots and affixes than those with no constituent morphemes. Likewise, Armon-Lotem and Chiat (in preparation) found significant effects of morphemes in both Russian (where wordlike items contained real affixes) and Hebrew (where wordlike items contained the root-pattern structure characteristic of real words in Semitic languages).

Relevance of factors in different languages

Human languages vary dramatically in the way they deploy the human phonetic capacity to create word phonologies, and the way phonology links with the semantic and syntactic properties of words. The phonological and morphological characteristics of words particular to a language determine potential influences on nonword repetition.

We might well assume that variations in word length are universal, but even this factor is not consistent across languages. In Cantonese, words are normally monosyllabic; most polysyllabic words are compounds, phrases, or English loan words (see Stokes et al., 2006). In English, most *monomorphemic* words are monosyllabic or bisyllabic, though polysyllabicity is common once words contain derivational morphemes. Thus, children's experience of word length varies according to their language. It is possible that this experience will influence the extent of length effects on nonword repetition, for example it may be that children who are exposed to words of three or more syllables from an early age will be better able to repeat longer nonwords than children whose early experience is largely confined words of one and two syllables. Indeed, this is put forward by Dispaldro et al. (2013) as a possible reason for the very high scores attained by typically developing children on their Italian nonword repetition test, and by Summers et al. (2010) as one possible reason why scores on their Spanish nonword repetition were higher than on their English nonword repetition test (as reported above). However, other phonological factors may be responsible for these cross-language differences, and it is possible that, all other things being equal, effects of length would override frequency of exposure to polysyllabic items in the target language. In order to address this issue, we would need to compare performance of children exposed to lexicons with opposite length biases, using tests that are the same on all variables apart from length. However, this theoretically interesting question may defy empirical investigation because length is conflated with other variables.

We have already seen that languages treat polysyllabic words in a variety of ways. In some, polysyllabic words are temporally like sequences of monosyllables, with even stress on each, and tone of syllables is contrastive; in some, the position of word stress is fixed, while others show complex patterns of stress. In stress-marking languages, some reduce unstressed vowels rendering unstressed syllables shorter than stressed syllables. Thus, increases in word length cannot be separated from prosodic structures which vary between languages. While we have some evidence that prosodic organisation of syllables facilitates repetition in English and Swedish (see above), these languages both favour trochaic patterns (in which unstressed syllables attach to a previous stressed syllable), making initial unstressed syllables atypical. We do not know how prosodic factors influence performance in typologically different languages, for example, if prosody has less or no effect in languages which do not reduce unstressed syllables. It may be that prosodic organisation of syllables is only helpful to children if this follows familiar prosodic patterns, but it is also possible that prosodic organisation facilitates recall regardless of children's particular language experience. This is an empirical question that COST Action IS0804 data will help to resolve.

Languages are similarly diverse in syllabic structure, varying in the number of consonants that may occur in onset and coda positions (i.e. before and after the vowel). Cantonese, for example, only allows singleton consonants, and the range of consonants that can occur in coda position is very limited. Finnish does not allow clusters, though some loan words contain clusters in onset position; it does, on the other hand, employ geminate (long) consonants. Hebrew allows clusters in onset but not coda position, and Turkish the reverse. Many languages, including Polish, Russian, English, German, and French, allow clusters of two or three consonants in one or both positions. We have extensive evidence that syllable complexity affects nonword repetition in languages that allow clusters, but we do not know whether children nevertheless benefit from familiarity with complex syllabic structures and are better able to repeat clusters if these occur in their language.

In addition to suprasegmental and syllabic differences, languages differ in their segmental repertoires, that is, in the numbers and types of consonant and vowel contrasts they deploy to create words. A few examples serve as a reminder of this point. Romance and Germanic languages share a range of fricative contrasts, but while German and Castilian Spanish, for example, use the voiceless fricative /x/, French and English do not; English and Castilian Spanish use /θ/, German and French do not. In contrast to English, which has around 20 vowels (depending on accent), Greek has just five. Even when languages employ a similar set of contrasts, for example voiced and voiceless plosives, the phonetic realisation of these contrasts is specific to the language, for example initial voiceless stops are aspirated in English but not in Spanish. The ways in which these segments combine with each other is also specific to the language. For example, French and English both allow word-initial /fr/, but only French allows [vr], and neither allow /pf/ which is legal in German. Finally, the frequency with which languages deploy segments and segmental combinations in their lexicon varies. Hence, phonotactic legality and probability are specific to a language.

The points made in this section are well known, and every reader will no doubt be able to think of segmental differences between languages with which they are acquainted and segments which are missing from their native repertoire and therefore difficult to produce. The purpose of drawing attention to these differences is to highlight the extent to which properties of nonwords known to influence nonword repetition are language-specific. However we construct nonwords, they will be more or less like real word phonology in different languages with respect to length, prosody, segmental constituents and their frequency, segmental combinations and their frequency, and phonetic realisation of segments. A particular length or prosody that is typical in one language may be disallowed or atypical in another; a syllable structure that is the norm in one language may be exceptional in another. Since nonwords necessarily carry these language-specific properties, a truly universal nonword test is not a possibility. This was an important issue for the development of the COST Action IS0804 nonword repetition framework described below.

Manipulation of factors in nonword repetition tests

Nonword repetition tests vary in the methods used to construct nonwords and the factors that are systematically manipulated. Length appears to be the only factor manipulated in all tests. Predictably, the widest range of tests and factors investigated is found in English. Indeed, the two most widely used English tests, the Children's Test of Nonword Repetition (CNRep; Gathercole & Baddeley, 1996) and the Nonword Repetition Test (NRT; Dollaghan & Campbell, 1998) lie at opposite ends of the spectrum with regard to a number of these factors. Both manipulate length of items, but beyond this, they diverge. Items in the CNRep incorporate many real English morphemes (e.g. *stoppograttic* contains the real word *stop* and adjectival suffix -*ic*). All items carry prosodic structure that is typical for English words, and prosody is not controlled. Items vary in segmental complexity, with many containing clusters (e.g. *skiticult*), and again, this factor is not controlled. Nor is perceptual or articulatory complexity of segments considered. In contrast, items in the NRT deliberately exclude any syllables that are real words or morphemes in order to eliminate any possible benefit from lexical knowledge. All syllables in NRT items contain tense vowels, and are produced with even stress, for example /dɔɪtɑʊvæb/, /tævɑtʃinɑɪg/. Their prosody makes items in the NRT quite different from real words in English, which are characterised by alternation of stressed and unstressed syllables, with unstressed syllables undergoing vowel reduction. In addition, NRT items are made up of simple CVCV structures, with no clusters, and do not include segments that are known to be challenging for children and emerge relatively late. In a further simplified test which aimed to eliminate any confounds from difficulties with speech sounds in nonwords, Shriberg et al. (2009) created a set of items that contained only four of the earliest occurring consonants and one early occurring vowel. Below, we consider how these differences between English nonword repetition tests influence their power to discriminate children with and without LI.

Few studies have pitted different factors against one another to determine whether these have independent or convergent effects (though effects of length, complexity and phonotactic probability are considered by Jones et al. (2010); effects of length and prosodic structure by Roy & Chiat, 2004, Chiat & Roy 2007; and effects of prosody and complexity by Marshall &

van der Lely, 2009). It would be difficult, if not impossible, to manipulate all relevant factors in a single test, since some factors cannot be disentangled from others. For example, more complex segmental structures are likely to be of lower phonotactic probability, and/or more difficult to perceive or articulate. However, manipulation of factors through test design is not the only way to investigate their relative effects. The Polish team in the COST Action IS0804 project (Szewczyk and Wodniecka, see Szewczyk, 2012) are using a different methodology to address this issue. Rather than manipulating selected factors to create Polish and English nonword repetition tests, Szewczyk computer-generated nonwords from real words, yielding items that vary on multiple parameters. Regression analyses can then be used to determine which factors make an independent contribution to children's performance, and the extent of that contribution. A key aim of this investigation is to throw more light on the skills that underpin nonword repetition and the role these play for different groups of children, monolingual and bilingual, with and without SLI.

Other teams within the COST Action IS0804 project have started from theories of phonological complexity, and used theoretically motivated parameters to drive the construction of nonwords, taking into account cross-linguistic differences that have emerged through the project. In particular, the French and German teams (Ferré & dos Santos, 2012; Grimm, 2012) have collaborated in developing tests that focus on the segmental content of nonwords, systematically varying complexity in terms of segments (more vs less difficult to produce) and syllable structure (with vs without clusters). In line with the broader COST Action IS0804 framework, these factors are manipulated in two sets of items, one designed to be maximally independent of the target language, and the other more language-dependent due to incorporation of language-specific features. Tests following these principles have been developed in collaboration with dos Santos and Ferré for Lebanese Arabic (Abi-Aad et al., 2013), Luxembourgish (Engel & dos Santos, 2012), and Serbian (Vuksanovic et al., 2013).

Other COST Action IS0804 researchers have developed tests controlled for one or more of the factors discussed above. These include Hebrew (Armon-Lotem et al., in preparation), Lithuanian (Dabašinskienė & Krivickaitė, 2012; Krivickaitė & Dabašinskienė, 2013), Slovakian (Kapalková et al., 2013), and Turkish (Topbaş & Kaçar, 2013).

The framework presented in this chapter (see next section) was designed to be compatible with the full range of language typologies represented in the group, with manipulation of universally applicable and language-specific factors. This framework is currently being applied in Finnish (Kunnari, personal communication), Greek-Cypriot (Kambanaros, 2011), Maltese and Maltese English (Grech & Calleja, personal communication), Serbian (Bjekic & Vuksanovic, 2012), English (Chiat, this chapter) and Irish English (Antonijevic-Elliott, personal communication). While typologically wide-ranging, these languages are far from fully representative of phonologies worldwide. We may therefore have overlooked factors that need to be considered in languages that are not represented in the working group, and exaggerated the importance of factors that are.

Relative effects of factors on children with LI and bilingual children

The variety of factors manipulated in nonword repetition tests begs questions about what type of nonword items best differentiate children with LI from TD children. It was important to take account of research addressing this question in creating our nonword repetition framework. Unsurprisingly, given the variety of nonword repetition tests and plethora of research studies in English, most of the evidence to date comes from English. In drawing on this evidence, it must be emphasised that outcomes may be specific to English language typology.

Findings on nonword repetition in English-speaking children with vs without LI have been reviewed in a number of papers (Gathercole, 2006; Coady & Evans, 2008). A meta-analysis conducted by Graf-Estes et al. (2007), collating results of 23 studies employing four different nonword repetition tests, is particularly pertinent. This meta-analysis revealed that the performance of SLI groups was on average 1.27 standard deviations below that of typically developing groups, and this did not change with age. However, the gap between TD and SLI groups varied considerably between tests. The most discriminating test was the CNRep (average effect size 1.94), and the least discriminating was the NRT (average effect size 0.9). This outcome is in line with findings of a study in which Archibald and Gathercole (2006) directly compared performance on these two tests in a group of 7-11 year olds. Both groups of children gained higher scores for percentage phonemes correct on the CNRep than on the NRT, but the CNRep produced a greater *discrepancy* between groups. Indeed, when children's *cognitive* ability was controlled, the group difference disappeared for the NRT, but was still present for the CNRep. We might infer that the CNRep taps into more language-specific skills and knowledge than the NRT (which makes sense when we consider the different characteristics of the two tests outlined above). The test that relies more on knowledge of the language is more challenging for children with language deficits.

As pointed out, the CNRep and NRT differ on a range of relevant parameters, so the greater deficit observed in the CNRep could arise from one or more of the parameters that differentiate it from the NRT. Many studies have examined the relative effect of particular parameters on children with vs without SLI. Interestingly, the prosodic factor on its own seems to produce a pattern of results similar to that observed for the CNRep vs NRT. In Archibald and Gathercole's (2007) study comparing repetition of syllables that were produced either as a list or as a single nonword (see above), the single nonword condition yielded higher scores in both groups of children (with and without SLI), but the difference between groups was greater in this condition, i.e. the SLI group showed a larger deficit in the single nonword than the list condition. It seems that both groups of children gain from having syllables prosodically organised in a single unit, but TD children benefit more than those with SLI. This is an interesting finding, and one which might throw light on the unique findings on nonword repetition in Cantonese. Stokes et al. (2006) point out that Cantonese lacks the prosodic and phonotactic complexity of languages such as English and Swedish. The Cantonese nonwords are therefore more like nonwords in the English NRT, but with the difference that they are like real polysyllabic words in Cantonese (characterised by even stress on syllables); this presumably makes them more familiar to Cantonese speakers compared with the NRT items whose prosody is uncharacteristic of English. It is also possible that the characteristics of the Cantonese nonwords make repetition more dependent on cognitive abilities, less dependent on phonological abilities, and therefore less vulnerable to deficits in phonological processing, as was found for the NRT (see Stokes et al., 2006, for further evidence).

When we look at the effects of other parameters on performance in children with and without SLI, the picture is mixed. Each parameter has been found to affect both groups. However, findings on the *relative effects* on children with and without SLI vary between factors and studies. In most cases, factors that make nonwords more difficult for TD groups make them disproportionately difficult for SLI groups. Thus, children with SLI have a greater deficit on items containing clusters (Jones et al., 2010; Archibald & Gathercole, 2006), and on longer items (Graf-Estes et al., 2007), though findings may vary according to the range of item lengths that are compared. Children with SLI have also been found to have a greater deficit on items with low phonotactic probability (Munson et al., 2005), but findings are not consistent (Jones et al., 2010).

If monolingual children with SLI are disproportionately affected by particular characteristics of nonwords such as length or syllable complexity, an important question is how these characteristics affect nonword repetition in bilingual children. Thordardottir and Brandeker (2013), for example, found that nonword length had particular effects on performance in their monolingual language-impaired group, but not in their typically developing bilingual group.

Evidence of how relevant factors affect language-impaired vs bilingual groups is currently very limited. This question merits further investigation: if bilingual children are less affected by length and/or syllable complexity than those with language impairment, these factors may help with clinical diagnosis in bilingual children when their overall scores fall below those of monolingual children.

The COST Action IS0804 nonword repetition framework

We have seen that items are more effective for detecting LI (in English) if they are prosodically structured, are longer, have more complex segmental structures, and perhaps if they contain lower frequency phoneme sequences. As pointed out above, these factors play out in different ways in the lexical structure of different languages. It is therefore possible that their effects on children with and without LI may vary according to the properties of real words that children have experienced. This encapsulates the challenges of constructing cross-linguistic nonword repetition tests. The quasi-universal and language-specific components that make up the COST Action IS0804 nonword repetition framework represent different reconciliations between the competing demands of maximum applicability across languages and maximum discrimination between children with and without SLI. They will enable us to evaluate the potential contribution of each type of test to the diagnosis of LI in bilingual children and children with limited experience of the target language.

Quasi-universal (QU) tests

We have already established the impossibility of a universal nonword repetition test. We have furthermore established that nonword repetition tests are informative because they tap children's phonological processing and representations, which are language-specific, and that tests and items containing more complex phonology are more informative about skills and deficits. Nevertheless, the COST Action IS0804 project provides a unique opportunity to run a set of almost identical nonwords across languages and find out if such a test discriminates TD and LI groups across languages, and how this compares with outcomes on language-specific tests. This was the rationale for creating the quasi-universal tests. Items in these tests were designed to be compatible with the cross-linguistically diverse constraints on lexical phonology discussed above. We designate the items *quasi*-universal because they necessarily have some language-specific characteristics, and even their shared properties are necessarily more or less characteristic of different languages.

The test contains 16 items, with equal numbers at 2, 3, 4 and 5 syllables. The items contain a limited range of consonants /p, b, t, d, k, g, s, z, l, m, n/ and vowels /a, i, u/, combined into simple CVCV structures. This makes them compatible with word phonology in most languages regardless of the further segmental contrasts and syllable structures that particular languages allow. Since any particular sequence of consonants and vowels may be a real word in a particular language, or contain a real word inflection, the test offers a set of options for each item. These are shown in Appendix 1. Once selected, the consonants and vowels in these items have to be produced with particular phonetic qualities, and these necessarily vary between languages. In the QU tests, consonants and vowels are assigned phonetic realisations that are characteristic of the dominant language in the child's environment (which may of course be distinct from the child's first language). In this respect, then, the items are language-specific.

QU test with quasi-neutral prosody: We have established that prosody varies between languages. This means it is not possible to assign a language-neutral prosody to QU items. However, we might assume that the most neutral prosody will be one that avoids any particular prosodic pattern by stressing all syllables equally. This reduces the possible influence of language-specific prosodic knowledge, and is the option we selected for the *QU test with quasi-neutral prosody*. Syllables in these items are produced with even length and pitch, apart from final syllable lengthening which characteristically marks the end of an utterance. It should be clear, though, that this even prosody is not truly neutral between languages: while it is close to the typical prosody for polysyllabic words in a language such as Cantonese (though lacking the tones that characterise Cantonese words), it is remote from prosody in languages such as English that reduce unstressed syllables, typically reducing the vowel to schwa.

A second QU test, the *QU test with language-specific prosody*, presents the same 16 items but with the prosodic pattern most typical for each word length in the target language. As pointed out above, we already have some evidence that lexical prosody benefits nonword repetition, and that nonwords with lexical prosody may yield greater differences between children with and without SLI (Archibald & Gathercole, 2007). However, this evidence is confined to English-speaking children, so we don't know whether these advantages are specific to languages in which polysyllabic words carry word-level prosodic patterns which distinguish them from a sequence of monosyllables, or whether they arise from universal language-processing mechanisms that make it easier to process phonological material when it is 'chunked' by prosody. Comparing performance on the two QU tests (with quasi-neutral vs language-specific prosody) in children exposed to languages that differ in their prosodic properties will help to address this question.

The English QU tests with quasi-neutral and language-specific prosody (Chiat, Polišenská & Szewczyk, 2012) are presented in Appendix 2.

Language-specific (LS) tests

The purpose of the language-specific tests within the COST Action IS0804 framework is to allow manipulation of phonological properties previously found to affect performance and relevant to particular languages. The LS framework includes 24 items, controlled for length, prosody, and syllable structure. In order to allow investigation of these factors without making the test excessively long, prosody and syllable structure are manipulated separately, but not in combination. Broad specifications for each parameter are proposed, with flexibility to cater for properties of lexical items in the target language. Where metrics for phonotactic probability are available for the target language, this factor is also manipulated. Wordlikeness ratings, on the other hand, can be obtained for all languages to provide a broad measure of their phonological familiarity.

Items in LS tests are consistent with constraints on lexical phonology in the target language. They are constructed from a representative range of consonants and vowels that occur in the language, avoiding only late-emerging consonants that may pose articulatory challenges to some children, and deploy these in phonotactically legal combinations. To minimise the potential contribution of lexical knowledge, syllables that are real morphemes in the target language should be avoided. The constituent segments are phonetically realised in accordance with the target language.

Length: The 24 items in the LS test are equally divided between lengths of 2-, 3- and 4-syllables. Monosyllabic items are not generally included since they are likely to produce ceiling effects, and add little to the evidence provided by bisyllabic items (Graf Estes et al., 2007). However, they should be included if real words in the language are typically monosyllabic. Conversely, 5-syllable items are not generally included because they are likely to produce floor effects in young children, and add little to evidence provided by 4-syllable items, particularly if a language allows complex syllabic structures. However, 5-syllable items should be included in languages where most lexical items are polysyllabic and 5-syllable items are common in children's vocabulary, and in those which lack the additional complexity of consonant clusters.

Prosody: In languages with *variable* word stress, six items with simple CV structure (see segmental complexity below) are assigned atypical stress patterns, two at each length. The remaining 18 items bear typical stress patterns.

Segmental complexity: In languages that permit consonant clusters, items with typical stress have no cluster or have one initial/medial cluster (depending whether one or both of these positions tolerate clusters in the target language), with equal numbers of each. Given the different possible combinations of factors in different languages, numbers of items with clusters will vary. The English LS test (Chiat et al., 2012) contains 18 items with typical stress patterns, six at each length, and these are equally divided between items with no cluster, initial cluster, and medial cluster, with two of each (see Appendix 2). In languages with fixed word stress, manipulation of syllable structure applies to all 24 items, allowing for a wider representation of clusters. These include eight at each length, divided as evenly as possible between no cluster, initial cluster, and medial cluster (according to positions in which clusters can occur).

In order to limit the number of factors and hence the length of tests, clusters are restricted to two consonants, but these should include several exemplars.

Phonotactic probability: Where metrics for phonotactic probability are available, subsets of items defined by length and, where applicable, prosody/segmental complexity, should be equally divided between high and low phonotactic probability, resulting in 12 of each.

Other lexical phonological factors: The factors specified so far reflect the lexical phonology of the languages most studied to date, with English obviously the most dominant. Lexical phonological parameters relevant to other languages will influence phonological familiarity and wordlikeness of nonwords, and should be taken into account in the construction of LS tests. For example, Hebrew and Arabic words are made up of *roots* (consonant frames) which are filled by *patterns* (vowels); in these languages, nonwords might be created from real root with non-pattern, non-root with real pattern, or non-root with non-pattern (Armon-Lotem et al., in preparation). Turkish and Finnish words are characterised by vowel harmony, and nonwords might be created with vs without harmony. In languages which inflect words according to their grammatical category, for example those that mark gender and/or case on nouns, nonword repetition tests may manipulate this factor, including items with and without recognisable inflections. However, it should be borne in mind that inclusion of real roots and patterns, vowel harmony, or inflections makes items more wordlike and therefore more susceptible to the influence of knowledge and experience.

Wordlikeness: Wordlikeness of nonwords is a subjective judgement of their similarity to real words. As pointed out, we might expect this to relate to the objective factors considered above, since native speakers presumably draw on their unconscious knowledge of typical prosodic and segmental patterns when they rate wordlikeness. Nonetheless, as a subjective judgement of similarity to real words, whatever the phonological and lexical factors influencing the judgement, wordlikeness is the one measure that may be truly cross-linguistic. For this reason, as well as the lack of lexicality metrics in some languages, it is useful to obtain wordlikeness judgements. To do this, all nonwords are entered on a sheet for native speakers to rate using a 5-point rating scale, with 5 most wordlike and 1 least wordlike. There are two options for the order of items on the sheet: the full set of QU and LS items may be randomized and presented in a single block, or QU nonwords with and without prosody and LS items may be randomised in three separate blocks, in which case order of the three blocks is counterbalanced across raters.

Order of testing

As with order of presentation for wordlikeness ratings, the COST Action IS0804 framework offers two options for order of presentation of items to children. All items may be randomised in one single set. In this case, the resulting randomised set may still be broken into several blocks to allow for breaks if children require these. The alternative is to randomise QU items with quasi-neutral prosody, QU items with language-specific prosody, and LS items in three separate blocks. In this case, the order of blocks should again be counterbalanced. Running the tests in these different ways will allow us to determine whether the effects of different types of item (QU with quasi-neutral/language-specific prosody and LS) are influenced by the properties of surrounding items. This will help to decide the order of items when the test format is finalised for clinical use.

Administration of tests

Items are recorded and recordings are incorporated into a PowerPoint presentation. COST Action IS0804 teams have developed a variety of child-friendly formats. The Polish test uses aliens to introduce nonwords (created by COST Action IS0804 member Engel de Abreu); the Finnish test (Kunnari) displays a set of stepping stones for each of the three tests, with one nonword attached to each stepping stone; the Turkish team (Topbaş and colleagues) use an animated parrot to deliver each nonword. The English test is available in two formats: the 'alien' format used in the Polish test, and a 'bead game' format developed by Polišenská and Kapalková (in press).

Whatever the format used, each item is played once, but a re-play is allowed if the first recording is interrupted, if the child is momentarily distracted, and for the first few items if the child gives no response to these. However, re-plays are not provided in response to children requesting another opportunity to hear the item again or improve on their first attempt.

Scoring

The COST Action IS0804 framework allows two levels of scoring: number of whole items correct and number of segments correct. Whole-item scoring is used in the CNRep and in the Preschool Repetition Test (Seeff-Gabriel, Chiat, & Roy, 2008). This has been found to be sufficiently discriminating, and is quicker and therefore more clinically practical (see also Dispaldro et al., 2013). Segmental scoring is used in the NRT and in some experimental studies using the CNRep. It is more time-consuming, but provides a more fine-grained measure of performance and may be more discriminating.

In the COST Action IS0804 framework, whole items are correct if they contain all and only the segments in the target in the correct order. Segments are correct if they fall within the target segmental category, even if they are phonetically distorted. Allowances are made for segmental substitutions that are relatively consistent in the child's productions and are characteristic of immature speech, for example, stopping of fricatives, fronting of velar stops in languages where these are observed in young children.

Conclusions and next steps

COST Action IS0804 work on the development of nonword repetition tests has thrown a spotlight on the diverse ways in which languages harness the human speech capacity to create words, evidenced by the rich cross-linguistic variation in lexical structure which must be considered when constructing nonwords. We have selected phonological properties that are widely applicable across languages for creation of our QU tests, and for LS tests, we have identified a range of parameters to be manipulated depending on the lexical typology of the target language. While motivated by evidence of cross-linguistic differences in lexical phonology and how these affect children's nonword repetition, our selection of properties and parameters is not necessarily optimal. It may be that the application of more abstract principles, with parametric variations appropriate to individual languages, would yield a more selective and informative set of items (dos Santos, Ferré and Grimm). Another possibility is that more comprehensive and sophisticated analyses of children's nonword repetition will reveal which factors contribute most to their performance, and/or which differentiate best between children with and without SLI (Szewczyk). Outcomes of such research might guide us towards reductions in the number of parameters to be manipulated in nonword repetition tests, leading to maximally short and informative tests.

Research using the COST Action IS0804 framework will also contribute new insights, helping to address a number of important empirical and theoretical questions. Cross-linguistic comparison of the two QU tests will clarify whether the parameters manipulated within and between these tests have a similar influence on children's performance regardless of their language background, or whether their influence varies according to the way these parameters play out in the language(s) to which children have been exposed. Together with findings from LS tests, they will also help to determine what types of nonwords are most informative about children's language abilities regardless of their language background.

Some caution is needed in making these cross-linguistic comparisons. Our nonword items are not identical across languages even in the QU tests. Different teams will be using different recordings of nonwords and different formats for presenting these, administering tasks in different combinations and different settings, and scoring them independently. Hence, any cross-linguistic differences we observe could be due to differences in items and/or methods rather than influences of language background. Nonetheless, if we find that certain factors have consistent effects across language groups, including languages where these factors play out differently, the evidence for universality of these factors will be strengthened, with implications for the phonological processing that nonword repetition assesses. Conversely, if we find that certain factors have variable effects across languages, and these are consistent with the role played by these factors in the lexical phonology of the language, this will strengthen the evidence for language-specific contributions to nonword repetition. The most definitive evidence will come from collaborative studies employing the *same recordings* of the QU tests (e.g. those made for English or Finnish) with different language groups, such that the pronunciation of consonants and vowels in both tests, and the prosody in the QU test with language-specific prosody, will be alien to at least one group. In such comparisons, any cross-group similarities and differences cannot be due to materials or methods, and must be attributed to linguistic experience and/or processing.

Finally, research on children with LI, both monolingual and bilingual, will throw more light on the extent to which the factors manipulated in the QU and LS tests affect their repetition of nonwords relative to typically developing peers, and which yield the most marked differences between groups. This will pave the way to identifying optimal nonword tests for our purposes: tests that yield the least difference between monolingual and bilingual children, and the greatest difference between children with and without LI.

REFERENCES

Abi-Aad, K., Atallah, C., Ferré, S. and dos Santos, C. (2013). Lebanese study: BiSLI and biTD. In Chiat, S., Ferré, S., dos Santos, C. and Grimm, A. Nonword repetition tests for assessing bilingual children. Paper presented at COST Action IS0804 Final conference Krakow 27-29 May 2013.

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

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

Armon-Lotem, S., Meir, N. and Chiat, S. (in preparation). The role of language experience in a STM task: Evidence from nonword repetition in Russian and Hebrew monolingual and sequential bilingual children.

Bishop, D.V.M., North, T. and Donlan, C. (1996). Nonword repetition as a behavioural marker for inherited language impairment: Evidence from a twin study. *Journal of Child Psychology and Psychiatry*, *37*, 391-403.

Bjekic, J., and Vuksanovic, J. (2012). Application of two quasi-universal nonword repetition tasks in Serbian SLI and TD children: A pilot study. Paper presented at COST Action IS0804 meeting held in Padova, 19-21 September 2012.

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

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

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

Chiat, S. and Roy, P. (2007).The Preschool Repetition Test: An evaluation of performance in typically developing and clinically referred children. *Journal of Speech, Language, and Hearing Research, 50,* 429-443.

Coady, J.A. and Evans, J.L. (2008).Uses and interpretations of non-word repetition tasks in children with and without language impairments (SLI).*International Journal of Language and Communication Disorders, 43,* 1-40.

Conti-Ramsden, G., Botting, N. and Faragher, B. (2001). Psycholinguistic markers for specific language impairment (SLI). *Journal of Child Psychology and Psychiatry, 42,* 741-748

Dabašinskienė, I. and Krivickaitė, E. (2012). Paper presented at COST Action IS0804 meeting held in Berlin, May 2012.

De Bree, E., Rispens, J. and Gerrits, E. (2007). Non-word repetition in Dutch children with (a risk of) dyslexia and SLI. *Clinical Linguistics & Phonetics, 21*, 935-944.

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

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

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

Engel, P. and dos Santos, C. (2012). Language bias in nonword repetition – towards a culture fair nonword repetition test. Paper presented at COST Action IS0804 meeting held in Berlin, 14-16 May 2012.

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

Engel, P.M.J., Santos, F.H. and Gathercole, S.E. (2008). Are working memory measures free of socioeconomic influence? *Journal of Speech, Language, and Hearing Research, 51*, 1580-187.

Ferré, S., and dos Santos, C. (2012). French NWRT: Inter-group comparison for French and Lebanese children. Paper presented at COST Action IS0804 meeting held in Padova, 19-21 September 2012.

Gathercole, S. E. (1995). Is nonword repetition a test of phonological memory or long-term knowledge? It all depends on the nonwords. *Memory and Cognition, 23,* 83-94.

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

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

Girbau, D. and Schwartz, R.G. (2007). Non-word repetition in Spanish-speaking children with Specific Language Impairment (SLI). *International Journal of Language and Communication Disorders, 42,* 59-75.

Graf Estes, K., Evans, J. and 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.

Grimm, A. (2012). The French-German NRT: How do German MON-SLI differ from eL2-SLI? Paper presented at COST Action IS0804 meeting held in 19-21 Padova, September 2012.

Gutiérrez-Clellen, V.F. 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,* 48–58.

Jones, G., Tamburelli, M., Watson, S.E., Gobet, F. and 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.

Kambanaros, M. (2011). Results of Greek Cypriot/Standard Greek test. Paper presented at COST Action IS0804 meeting held in Malta, 28-30 November 2011.

Kapalková, S., Polišenská, K. and Vicenová, Z. (2013). Non-word repetition performance in Slovak-speaking children with and without SLI: Novel scoring methods. *International Journal of Language and Communication Disorders*, *48,* 78-89.

Kavitskaya, D., Babyonyshev, M., Walls, T. and Grigorenko, E. (2011). Investigating the effects of syllable complexity in Russian-speaking children with SLI. *Journal of Child Language, 38,* 979-998.

Kohnert, K., Windsor, J. and Yim, D. (2006). Do Language-Based Processing Tasks Separate Children with Language Impairment from Typical Bilinguals? *Learning Disabilities Research & Practice, 21,* 19–29.

Krivickaitė, E. and Dabašinskienė, I. (2013). Lithuanian nonword repetition test. Paper presented at COST Action IS0804 meeting held in Lisbon, 13-15 February 2013.

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

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

Marshall, C. R. and van der Lely, H. K. J. (2009). Effects of word position and stress on onset cluster production: Evidence from typical development, SLI and dyslexia. *Language, 85,* 39-57.

Messer, M.H., Leseman, P.P.M., Boom, J. and 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.

Munson, B. (2001). Phonological pattern frequency and speech production in children and adults.*Journal of Speech, Language, and Hearing Research*, 44, 778-792.

Munson, B., Edwards, J. and Beckman, M.E. (2005).Phonological knowledge in typical and atypical speech-sound development. *Topics in Language Disorders, 25,* 190-206.

Munson, B., Kurtz, B.A. and Windsor, J. (2005).The influence of vocabulary size, phonotactic probability, and wordlikeness on nonword repetitions of children with and without specific language impairment. *Journal of Speech, Language, and Hearing Research, 48,* 1033-1047.

Polišenská, K., Chiat, S., and Szewczyk, J. (in preparation). An investigation of relations between objective properties of nonwords and subjective judgements of their wordlikeness.

Polišenská, K. and Kapalkova, S. (in press). Improving child compliance on a computer administered nonword repetition task. *Journal of Speech, Language, and Hearing Research*.

Roy, P. and 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,* 223-34.

Roy, P. and 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. and Radeborg, K. (1999). Nonword repetition in children with language impairment - pitfalls and possibilities. *International Journal of Language and Communication Disorders, 34,* 337-352.

Seeff-Gabriel, B., Chiat, S. and Roy, P. (2008).*The Early Repetition Battery*. London: Pearson Assessment.

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

Shriberg, L.D., Lohmeier, H.L., Campbell, T.F., Dollaghan, C.A., Green, J.R. and Moore, C.A. (2009). A nonword repetition task for speakers with misarticulations: The Syllable Repetition Test (SRT). *Journal of Speech, Language, and Hearing Research, 52,* 1189-1212.

Stokes, S.F., Wong, A., Fletcher, P. and Leonard, L.B. (2006). Nonword repetition and sentence repetition as clinical markers of specific language impairment: The case of Cantonese. *Journal of Speech, Language, and Hearing Research, 49,* 219-236.

Szewczyk, J. (2012). Subject- and item-related predictors of nonword repetition performance, based on the Polish NWR test. Paper presented at COST Action IS0804 meeting held in Berlin, May 2012.

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

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,* 922–937.

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,* 1–16.

Thordardottir, E.T. and 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.

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

Vuksanovic, J., Bjekic, J. and Zivanovic, M. (2013). Application of two nonword repetition tasks in TD Bilingual Serbian-Hungarian children. Paper presented at COST Action IS0804 meeting held in Lisbon, 13-15 February 2013.

Williams, D., Payne, H. and Marshall, C. R. (2013). Non-word repetition impairment in autism and SLI: Evidence for distinct underlying cognitive causes. *Journal of Autism and Developmental Disorders, 43,* 404-417.

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

Appendix 1: Construction of quasi-universal tests

<table>
<tr><td colspan="7">• For each item (1-16), select any one example in the row which is not a real word in the target language:</td></tr>
<tr><td>1</td><td>zibu</td><td>sibu</td><td>sipu</td><td>zipu</td><td></td><td></td></tr>
<tr><td>2</td><td>lita</td><td>lida</td><td>dula</td><td>tula</td><td></td><td></td></tr>
<tr><td>3</td><td>maki</td><td>naki</td><td>magi</td><td>nagi</td><td></td><td></td></tr>
<tr><td>4</td><td>luni</td><td>lumi</td><td>nuli</td><td>muli</td><td></td><td></td></tr>
<tr><td>5</td><td>sipula</td><td>zipula</td><td>sibula</td><td>zibula</td><td></td><td></td></tr>
<tr><td>6</td><td>bamudi</td><td>banudi</td><td>pamudi</td><td>panudi</td><td>pamuti</td><td>panuti</td></tr>
<tr><td>7</td><td>malitu</td><td>malidu</td><td>nalitu</td><td>nalidu</td><td>malitu</td><td>malidu</td></tr>
<tr><td>8</td><td>lumika</td><td>lunika</td><td>lumiga</td><td>luniga</td><td></td><td></td></tr>
<tr><td>9</td><td>zipalita</td><td>sipalita</td><td>zibalita</td><td>sibalita</td><td>zipalida</td><td>sipalida</td></tr>
<tr><td>10</td><td>mukitala</td><td>nukitala</td><td>mugitala</td><td>nugitala</td><td>mukidala</td><td>mugidala</td></tr>
<tr><td>11</td><td>kasulumi</td><td>gasulumi</td><td>kazulumi</td><td>gazulumi</td><td>kasuluni</td><td>gasuluni</td></tr>
<tr><td>12</td><td>litisaku</td><td>lidisaku</td><td>litisagu</td><td>lidisagu</td><td>litizaku</td><td>lidizaku</td></tr>
<tr><td>13</td><td>sipumakila</td><td>sibumakila</td><td>sipunakila</td><td>sibunakila</td><td>sipumagila</td><td>sibumagila</td></tr>
<tr><td>14</td><td>tulikasumu</td><td>dulikasumu</td><td>tuligasumu</td><td>duligasumu</td><td>tulikazumu</td><td>dulikasumu</td></tr>
<tr><td>15</td><td>malusikuba</td><td>maluzikuba</td><td>malusiguba</td><td>maluziguba</td><td>malusikupa</td><td>maluzikupa</td></tr>
<tr><td>16</td><td>litapimuti</td><td>lidapimuti</td><td>litabimuti</td><td>lidabimuti</td><td>litapimudi</td><td>lidabimudi</td></tr>
<tr><td colspan="7">• Pronounce each item with vowel and consonant qualities appropriate to the target language.</td></tr>
<tr><td colspan="7">• QU with quasi-neutral prosody: Produce items with even stress on each syllable, and if appropriate for target language, final lengthening on final syllable.</td></tr>
<tr><td colspan="7">• QU with language-specific prosody: Produce with the prosodic pattern most typical for each word length in the target language.</td></tr>
</table>

Appendix 2: Quasi-universal tests for English (Chiat, Polišenská & Szewczyk, 2012)

With quasi-neutral prosody: **With English consonants and vowels**	**With language-specific prosody:** **With English consonants and vowels, and prosody typical for word length**
ˈ even stress and pitch ˌ falling pitch	ˈ primary stress ˌ secondary stress
/ˈziˌbu/ /ˈduˌlɑ/ /ˈnɑˌgi/ /ˈluˌmi/	/ˈzibə/ /ˈdulə/ /ˈnɑgi/ /ˈlumi/
/ˈsiˈpuˌlɑ/ /ˈbɑˈmuˌdi/ /ˈmɑˈliˌtu/ /ˈluˈmiˌgɑ/	/ˈsipəˌlɑ/ /ˈbɑməˌdi/ /ˈmɑlɪˌtu/ /ˈlumɪˌgɑ/
/ˈziˈpɑˈliˌdɑ/ /ˈmuˈkiˈtɑˌlɑ/ /ˈkɑˈsuˈluˌmi/ /ˈlidiˈsɑˌku/	/ˈzipəˌlidə/ /ˈmukɪˌtɑlə/ /ˌkɑsəˈlumi/ /ˌlidɪˈsɑkə/
/ˈsiˈpuˈmɑˈkiˌlɑ/ /ˈduˈliˈgɑˈsuˌmu/ /ˈmɑˈluˈziˈguˌbɑ/ /ˈliˈtɑˈpiˈmuˌti/	/ˌsipəmæˈkilə/ /ˌdulɪgæˈsumə/ /ˌmɑləˈzigəbə/ /ˌlitəˈpiməti/

Appendix 3: Language-specific test for English (Chiat, Polišenská & Szewczyk, 2012)

Stress	**Complexity**	**No Syll**	**Transitional probability (TP) / Ngram frequency (NF)[1]**			
			High		**Low**	
			Item no.	Item TP / NF	*Item no.*	Item TP / NF
Typical	- cluster	2	*1*	ˈdallen ˈdælən 11 / 4.56	*13*	ˈrefap ˈrɛfəp 8.8 / 3.12
		3	*2*	ˈsannery ˈsænəri 16.8 / 3.42	*14*	ˈzummerlah ˈzʊməlɑ 4.6 / 2
		4	*3*	ˌponnerˈvayker ˌpɒnəˈveɪkə 13.9 / 2.5	*15*	ˌkefferˈmoyper ˌkɛfəˈmɔɪpə 10.7 / 1.6
	+ initial cluster	2	*4*	ˈspoddle ˈspɒdəl 9.6 / 2.96	*16*	ˈfrashek ˈfræʃək 7.1 / 2.72
		3	*5*	ˈstoffely ˈstɒfəli 15.9 / 2.79	*17*	ˈsmisherˌtow ˈsmɪʃəˌtaʊ 9.5 / 2.28
		4	*6*	ˌskoomerˈkider ˌskuməˈkaɪdə 10.4 / 1.94	*18*	ˌflahnerˈmoozer ˌflɑnəˈmuzə 8.8 / 1.79
	+ medial cluster	2	*7*	ˈnahsket ˈnɑskət 12 / 3.15	*19*	ˈlursnok ˈlɜsnɒk 5.6 / 2.44
		3	*8*	ˈmahsperˌdow ˈmɑspəˌdaʊ 8.8 / 2.25	*20*	ˈzeespegoy ˈzispəˌgɔɪ 4.8 / 1.35
		4	*9*	ˌtoskerˈleemer ˌtɒskəˈlimə 8.7 / 2.1	*21*	ˌvosnaˈrouder ˌvɒsnəˈraʊdə 7.6 / 1.81
Atypical	- cluster	2	*10*	reˈvike rɪˈvaɪk 12 / 3.63	*22*	naˈlorsh nəˈlɔʃ 3 / 2.28
		3	*11*	peˈzayner pəˈzeɪnə 11.7 / 2.88	*23*	leˈvooger ləˈvugə 9.4 / 2.05
		4	*12*	reˈnusedar rəˈnusədɑ 5.3 / 1.96	*24*	zeˈdahgenur zəˈdɑgənɜ 3.6 / 1.49

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