Non-word repetition in adolescents with SLI

Susan H. Ebbels,
Moor House School, Surrey, UK and
Division of Psychology and Language Sciences, University College London, UK

Julie E. Dockrell
Psychology and Human Development, Institute of Education, London, UK

Heather K.J. van der Lely
Department of Psychology, Harvard University, Cambridge MS,
Laboratoire de Neuropsychologie Interventionnelle, Départament Etudes Cognitives,
École Normale Supérieur, Paris, France and
Psychological Sciences Research Institute, Université Catholique de Louvain, Louvain de Neuve, Belgium

Correspondence concerning this article should be addressed to Susan Ebbels, Moor House School, Mill Lane, Hurst Green, Oxted, Surrey, RH8 9AQ, UK. ebbelss@moorhouseschool.co.uk or to Heather van der Lely, Psychology Department, William James Hall, 33 Kirkland Street, Cambridge, MA 02138. hvdlely@wjh.harvard.edu

Keywords:

Specific language impairment (SLI), adolescents, non-word repetition, verb agreement and tense marking, phonological complexity

Abstract

Background: Non-word repetition (NWR) difficulties are common, but not universal, among children with Specific Language Impairment (SLI). However, older children and adolescents with SLI have rarely been studied. Studies disagree on the relationship between NWR difficulties and difficulties with other areas of language and literacy. There is also no consensus as to the underlying reason for the difficulties (some) children with SLI have with NWR. Some scholars argue that difficulties with phonological short-term memory or storage cause NWR and other language difficulties, whereas others argue that difficulties with NWR may be due more to difficulties with phonological representations.

Aims: To investigate NWR abilities and their relationship to other language and literacy abilities in a group of older children with SLI and typically developing (TD) controls. To investigate the relative effects of increasing phonological complexity and the number of syllables on the ability of the participants to repeat non-words.

Methods & Procedures: We administered a NWR test (The Test of Phonological Structure, TOPhS) which systematically varies phonological complexity to 15 participants with SLI (11-15 years), 30 language and 15 age controls. We also administered standardised language and literacy tests and a specific test of verb agreement and tense marking (Verb Agreement and Tense Test, VATT).

Outcomes & Results: The participants with SLI showed a bi-modal distribution: half achieved age appropriate NWR while half scored significantly below language and age controls (d>7). The two groups of participants with SLI (high versus low scorers) only differed in NWR (d>5) and agreement (d>3) and tense marking (d>2.5), not on our other standardised language and literacy measures. NWR was also highly correlated with verb agreement (r=.97) and tense marking (r=.89) among participants with SLI, but not among controls (r=.16 and r=.30

respectively). Phonological complexity was related to NWR accuracy, particularly among participants with SLI. The number of syllables had no independent effect on NWR performance for any group.

Conclusions & Implications: Some children with SLI (who have good NWR) have language difficulties unrelated to any of the factors underlying NWR. Others have a (probably additional) deficit which affects NWR and also leads to greater difficulties with verb agreement and tense marking. Our results indicate that difficulties with this particular NWR test are more likely to be due to a deficit with phonology per se, rather than phonological short-term memory or storage.

What this paper adds

What is already known on this subject

Many children with SLI find repeating non-words hard. It is unclear a) whether all children (particularly older children and adolescents) with SLI have NWR difficulties, b) how these difficulties are related to their general language and literacy difficulties and c) what underlies their NWR difficulties.

What this study adds

We found that only half of our adolescent participants with SLI had NWR difficulties, despite having severe and persistent difficulties with other areas of language. Thus, difficulties with NWR (or indeed any of the possible factors which underlie NWR difficulties) are unlikely to be the underlying cause of the language impairments of at least half of our participants. Difficulties with NWR were related to difficulties with verb agreement and tense marking but not to performance on general standardised language or literacy tests. A preliminary analysis of the factors in the non-words which may affect performance suggests that phonological complexity is more strongly related to performance than the number of syllables, particularly among children with SLI. This suggests difficulties with phonological representations are more likely to underlie their NWR difficulties than phonological short-term memory or storage, but this requires further investigation.

Introduction

Children with Specific Language Impairment (SLI) have difficulties acquiring language despite adequate intelligence, hearing, physical development and exposure to language. When compared with age controls, they score lower on most areas of language (e.g., vocabulary, lexical learning, morphology, syntax, argument structure, text comprehension). However, they also score lower than controls matched on general language abilities in specific areas. The areas of morpho-syntax and syntax have been well researched and a general consensus has been reached that (English) children with SLI have particular difficulties with verb tense and agreement, making more errors than language controls (e.g., Leonard, Bortolini, Caselli, McGregor, & Sabbadini, 1992; Rice, Wexler, & Cleave, 1995) as well as a number of areas of syntax (e.g., Bishop, 1979; van der Lely, 2005). In addition, a key finding over many years is of particular problems with non-word repetition (NWR) (e.g., Briscoe, Bishop, & Norbury, 2001; Edwards & Lahey, 1998; Farmer, 2000; Gallon, van der Lely, & Harris, 2007; Gathercole & Baddeley, 1990; Montgomery, 2004; Montgomery, 1995). The relationship between NWR difficulties and other core deficits is the subject of considerable discussion (see Gathercole, 2006).

A recent meta-analysis of NWR studies including children with SLI (Graf Estes, Evans, & Else-Quest, 2007) gave a combined effect size of *d*=1.27 (ranging from 0.62 to 4.34 in the studies analysed) for the difference between children with SLI and age controls. The effect sizes were not associated with the age of the children being tested, which suggests the magnitude of the deficit in SLI remains relatively stable across ages. Indeed differences between children with SLI and typically developing (TD) children have been found in older children and adolescents with SLI aged 8-13 years (Munson, Kurtz, & Windsor, 2005) and 12-19 years (Gallon et al., 2007).

Variation in NWR performance among children with SLI

NWR has been proposed as a culturally unbiased clinical marker for SLI which is independent of IQ or social economic status (e.g., Conti-Ramsden, Botting, & Faragher, 2001; Gardner, Froud, McClelland, & van der Lely, 2006). However, several studies have reported an overlap between the scores of children with SLI and those of TD children (Bishop, North, & Donlan, 1996; Edwards & Lahey, 1998; Gallon et al., 2007; Montgomery, 1995). This indicates that there are children with SLI who do not have difficulties repeating non-words. Indeed Botting and Conti-Ramsden (2001) reported that 147 out of 200 children with SLI scored below -1SD for their age on a NWR test. Therefore, by implication, 53 (26.5%) must have scored within or above the normal range. Some recent studies have documented the numbers or percentages of children with language impairments (or at risk of language impairments) who are not impaired on NWR tests. Chiat and Roy (2007) found 25% of their SLI group aged 2;6-4 years scored within the normal range (>-1SD) on their repetition test which included both words and non-words. Gardner et al. (2006) found 59% of children with SLI passed the Grammar and Phonology Screening (GAPS) phonology subtest (using a NWR procedure). Bishop, Adams and Norbury (2006) also found that of children who had difficulties marking verb agreement and tense accurately, less than half also had difficulty repeating non-words. These studies, however, only considered children under the age of twelve years. Variation in performance in older children with SLI to our knowledge has not yet been investigated. This is important to study as when children do not show difficulties with NWR, it is unlikely that they have ongoing difficulty with any of the lower level processes involved in this procedure (e.g., auditory or phonological processing, phonological storage or short term memory).

One factor which may be related to variation in SLI performance on NWR tasks is the presence or absence of co-occurring dyslexia as children with both SLI and dyslexia perform significantly worse on NWR tasks than children with SLI only (Baird, Slonims, Simonoff, &

Dworzynski, 2011; Bishop, McDonald, Bird, & Hayiou-Thomas, 2009; Catts, Adlof, Hogan, & Weismer, 2005; Rispens & Parigger, 2010). Indeed, two of these studies (Bishop et al., 2009; Rispens & Parigger, 2010) found that the children with SLI only did not differ from TD age controls.

Relationship between NWR and other language and literacy measures

Correlational evidence of a link between NWR and other language measures in SLI is inconclusive. Some studies have calculated correlations using a combined (SLI and TD) group. This assumes that children with SLI and TD children lie on a continuum. However, children with SLI may have specific difficulties which affect particular language components (e.g., van der Lely, 2005). Therefore grouping them together with TD children may mask potential differences in correlations between the two groups. For those studies which have considered children with SLI separately, NWR has been found to correlate significantly with expressive language as measured on general language tests (e.g., Edwards & Lahey, 1998; Montgomery & Windsor, 2007). Some studies found significant correlations with sentence comprehension (e.g., Montgomery, 1995; Sahlen, Reuterskiold-Wagner, Nettelbladt, & Radeborg, 1999), whereas others did not (e.g., Gallon et al., 2007; Montgomery, 2004). Botting and Conti-Ramsden (2001) found that NWR predicted all language measures apart from vocabulary. Indeed, significant correlations between NWR and vocabulary are rarely reported for children with SLI (e.g., Farmer, 2000; Hansson, Forsberg, Lofqvist, Maki-Torkko, & Sahlen, 2004), with more studies finding no significant correlation (e.g., Briscoe et al., 2001; Edwards & Lahey, 1998; Gallon et al., 2007). This is despite such a correlation being found in (generally younger) TD children (e.g., Briscoe et al., 2001; Gathercole & Baddeley, 1989). The only study we could find reporting correlations between literacy and NWR measures (de Bree, Wijnen, & Gerrits, 2010) found no correlation between these among children with SLI, but did find a correlation among children at risk of dyslexia.

For verb agreement and tense marking, one study reported weak, but non-significant correlations with NWR within their SLI group (Norbury, Bishop, & Briscoe, 2001) while others reported significant correlations (Bishop, Adams, & Norbury, 2006; Thordardottir, 2008). Formation of the regular past tense and also 3rd person singular agreement usually results in a consonant cluster and Marshall and van der Lely (2007) found that children with SLI (but not TD children) were less likely to produce past tense forms that require a word final cluster (a phonologically more complex structure) than those that don't. Thus, phonological abilities can impact on verb agreement and tense marking. Phonological abilities are also likely to impact on NWR performance and at least some of the variation in NWR in SLI discussed above could be due to variation in phonological abilities. Thus, a correlation would be predicted between a past tense and agreement test requiring clusters and a NWR test including clusters in children with SLI, because of the impact of phonological abilities on both tasks. However, if some children with SLI do not have difficulties with NWR, as suggested by previous studies, they presumably do not have phonological difficulties (or difficulties with any of the other processes underlying NWR). Then, the question remains as to whether they would have difficulties with verb agreement and tense marking, as such difficulties have been suggested as a marker for SLI (e.g., Rice & Wexler, 1996). Although phonology may have an impact on this area, knowledge of syntactic and morphological rules and the ability to apply them are also required and may remain impaired in all children with SLI, even in those whose phonological abilities may be relatively unimpaired.

Factors underlying difficulties with NWR

The underlying causes of the difficulties many children with SLI have with NWR and the extent to which these cause difficulties with other areas of language have been the subject of considerable debate (see Gathercole, 2006 and associated commentaries). Part of the reason for this lack of agreement is that NWR is a complex task which involves several cognitive processes (Snowling, Chiat, & Hulme, 1991), any of which could underlie difficulties with the task. Different NWR tasks are designed to test different properties of non-words. Gathercole and Baddeley's (1990) Children's Non-word Repetition (CNRep) test aimed to test phonological short-term memory and therefore varies the length of non-words in order to analyse the effect of length on performance. Dollaghan and Campbell's (1998) Nonword Repetition Test (NRT) also varies length, but was designed to minimise complexity, so excludes clusters and late-acquired consonants. In contrast, Edwards and Lahey (1998) manipulated the phonological complexity of their non-words and carried out detailed error analyses to rule out difficulties with discrimination or planning and executing a response. In this study, we used the Test of Phonological Structure (TOPhS, van der Lely & Harris, 1999), which systematically manipulates prosodic-phonological complexity rather than length, thus aiming to test phonological representations themselves rather than the associated processes of phonological short-term memory or storage.

Previous studies have mainly examined the effect of length (usually measured in number of syllables) and phonological complexity. Findings of a length effect have been used to support theories implicating phonological short-term memory (Gathercole & Baddeley, 1990) or storage (Gathercole, 2006) and findings of a phonological complexity effect to support theories implicating phonology and phonological representations (Chiat, 2001; Edwards & Lahey, 1998; Gallon et al., 2007; Marshall & van der Lely, 2009). Most studies which considered word length found performance decreased with increasing word length for all children, but particularly those with SLI (e.g., Bishop et al., 1996; Dollaghan & Campbell, 1998). In individual studies, differences between children with SLI and TD children at shorter word lengths were often not significant but when combined in a meta-analysis, Graf-Estes et al. (2007) found children with SLI were impaired at all word lengths, but effect size magnitudes increased with word length. Thus, while phonological short-term memory or storage could be involved, impairment on even short non-words implicates other mechanisms or processes.

One difficulty with analyses of length effects is that in many NWR tests (especially the CNRep), length is conflated with other factors such as phonological complexity (longer words often have more complex stress patterns, or ‘metrical structure’), phonotactic probabilities and possibly also wordlikeness (which is itself related to phonotactic probability, Munson et al., 2005). Munson et al. (2005) found a main effect of length using non-words containing either three or four syllables. But when they carried out a regression analysis including wordlikeness and phonotactic probability, length did not predict a significant proportion of the variance, unlike the other two factors.

Several studies have considered the impact of phonological complexity in terms of the presence or absence of consonant clusters. All have found that non-words involving consonant clusters are harder to repeat, but some have found this affects both children with SLI and TD children equally (Edwards & Lahey, 1998; Gathercole & Baddeley, 1990) while others have found a complexity by group interaction where children with SLI were more affected by the addition of consonant clusters than TD children not only in general (Archibald & Gathercole, 2006; Bishop et al., 1996; Briscoe et al., 2001) but also even on monosyllabic non-words (Gallon et al., 2007).

Another aspect of phonological complexity concerns words' metrical structure (i.e., stress patterns). Investigations of metrical structure have revealed that both young TD children and young children with SLI omit more syllables if they are unstressed and if they are in a pre-stress rather than post-stress position (Chiat & Roy, 2007). Omission of syllables is less frequent in older children with SLI (Gallon et al., 2007; Marshall & van der Lely, 2009), but these studies found metrical structure still impacts on the ability of children with SLI and those with dyslexia (but not TD children, Marshall & van der Lely, 2009) to repeat non-words..

Summary

NWR has been proposed as a possible clinical marker of SLI, but variability in performance has been reported for younger children with SLI. Only one study (Gallon et al., 2007) has included adolescents with SLI and thus the degree of variability in this age group is uncertain. We therefore investigated the NWR abilities of older children with SLI comparing them with both age and language controls, extending Gallon et al.'s (2007) study where they used several groups of language controls but not age controls.

NWR has been found to be related to other language measures in SLI, but this is not universal. We therefore investigated correlations between NWR and a range of general language and literacy measures and also a specific measure tapping verb agreement and tense marking, within both children with SLI and TD children. We predicted that the children with SLI would have particular difficulties with verb agreement and tense marking, as this has also been proposed as a clinical marker for SLI (e.g., Rice & Wexler, 1996). However, as Marshall and van der Lely (2007) pointed out, phonological complexity impacts on past tense and agreement. Therefore, this predicts that there is likely to be a significant correlation between the children's performance on verb agreement and tense marking and NWR (using the TOPhS) as both involve complex phonology (consonant clusters).

Furthermore, while there is a general consensus that NWR is constrained by multiple processes, the relative importance of these processes, particularly as concerns the poor NWR of children with SLI is still the topic of much debate (see Gathercole, 2006 and associated commentaries). Many researchers view poor phonological short term memory or storage as the root cause while others propose difficulties forming phonological representations could underlie poor NWR performance. We will investigate this using the TOPhS whose stimuli vary in both complexity and length. These variables are correlated, but not perfectly. We are therefore able to do a preliminary analysis by items to determine whether length and/or phonological complexity make an independent contribution to the variance in TOPhS scores.

Aims of study

1. Evaluate Gallon et al.'s (2007) finding that older children and adolescents with SLI have more difficulties with NWR than TD children and investigate variation in NWR performance among this older SLI group.
2. Examine correlations between NWR performance on the TOPhS and other language and literacy measures in children with SLI and TD children separately.
3. Investigate the relative contributions of length and phonological complexity to the ability of children with SLI and TD children to repeat non-words accurately.

Methods

Participants

Fifteen participants with persisting SLI (mean age: 13;1 years, range: 11;0–14;11), thirty language controls (mean age: 8;5, range: 5;4-12;2) and fifteen age controls (mean age: 13;1, range: 11;3-14;10) participated. The participants with SLI all attended a specialist school in the UK for children with primary language impairments. All children in the school who were aged between 11 and 15 and met the following criteria were recruited: 1) receptive and expressive language difficulties (at least –1.5 SD below the mean) as measured on the Clinical Evaluation of Language Fundamentals-3 (CELF-3 UK, Semel, Wiig, & Secord, 1995), 2) typical non-verbal performance abilities (not more than –1 SD below the mean) on the mean of Matrices and Pattern Construction from the British Ability Scales II (BAS-II, Elliot, Smith, & McCulloch, 1996), 3) intelligible spontaneous speech, 4) no hearing impairment, neurological dysfunction, structural abnormalities or diagnosis of autism or Asperger's syndrome, 5) written consent given by the parents.

The participants with SLI were also tested on the Wechsler Objective Reading Dimensions (WORD, Rust, Golombok, & Trickey, 1993). As a group, they scored on average in the impaired range on Reading Accuracy (mean SS: 72) and Spelling Accuracy (mean SS: 75). At an individual level, three participants scored within the normal range on both Reading and Spelling accuracy (all the others scored more than 1.5 standard deviations below the mean on Reading accuracy, although two of these scored within the normal range on Spelling accuracy). Thus, for only three of the participants with SLI could dyslexia be ruled out.

The TD controls were recruited from six mainstream schools in the same geographical region as the SLI school. None had identified special educational needs or English as an additional language. They scored above –1SD on both performance IQ (mean of Matrices and Pattern Construction from the BAS-II) and language abilities. Language was tested using the Formulated Sentences subtest of the CELF-3 (which requires participants to create sentences including particular given words), the British Picture Vocabulary Scale – II (BPVS-II, a multiple-choice vocabulary comprehension test, Dunn, Dunn, Whetton, & Burley, 1997) and the Test of Reception Of Grammar (TROG, a multiple-choice sentence comprehension test, Bishop, 1989). Each language control child was individually matched to a child with SLI on the basis of performance IQ (within one standard deviation) and either the BPVS (15 'BPVS controls': raw score within 3 points) or the TROG (15 'TROG controls': matched on exact raw score). They were also required to score within the normal range for their age (i.e., not more than 1 SD above or below the mean) on the test with which they were matched to the participants with SLI. The age controls were individually matched to the participants with SLI by age (within three months) and also scored within the normal range (within one standard deviation of the mean) on the BAS-II (performance IQ)[1]. The scores for the four groups on the matching criteria are shown in Table 1.

TABLE 1 ABOUT HERE

In order to validate the matching procedures, the groups were compared on age and raw scores of the language tests used for matching. We found a significant effect of age, $F(3,56)=51.18$, $p<0.001$, $\eta p^2=0.73$, where the participants with SLI did not differ from the Age controls, $p=1.0$, $d=0.004$, but differed significantly from both TROG, $p<0.001$, $d=3.35$ and BPVS controls, $p<0.001$, $d=3.03$. The latter two groups did not differ from each other, $p=1.0$, $d=0.3$, but did differ from the Age controls (TROG: $p<0.001$, $d=3.34$, BPVS: $p<0.001$, $d=3.03$).

The four groups differed significantly on the BPVS raw score, $F(3,56)=18.874$, $p<0.001$, $\eta p^2=0.50$. Post-hoc tests showed the participants with SLI did not differ from either their BPVS, $p=1.0$, $d=0.01$, or TROG controls, $p=1.0$, $d=0.28$, but scored significantly lower than their Age controls, $p<0.001$, $d=2.13$, as did both the TROG, $p<0.001$, $d=2.41$, and BPVS controls, $p<0.001$, $d=2.15$, who did not differ from each other, $p=1.0$, $d=0.27$.

The four groups also differed significantly on the TROG raw score, $\chi^2(3)=23.46$, $p<0.001$[2]. Post-hoc tests showed the participants with SLI did not differ from either their TROG, $W=232.5$, $n_1=15$, $n_2=15$, $p=1.0$ or BPVS controls, $W=74.5$, $n_1=15$, $n_2=15$, $p=0.11$, but did differ from their Age controls, $W=130$, $n_1=15$, $n_2=15$, $p<0.001$. The TROG controls differed from the Age controls, $W=130$, $n_1=15$, $n_2=15$, $p<0.001$, whereas the BPVS controls did not, $W=185.9$,

[1] One age control achieved a z-score of –1.15 on the BAS, but showed no language difficulties, was matched to the child with SLI with the lowest z-score (-0.95) on the BAS and was exactly the same age; he was therefore considered to provide a good match.

[2] The non-parametric Kruskal-Wallis and Wilcoxon signed ranks tests were used because the data for the SLI and language control groups were not normally distributed (the SLI and TROG control groups were positively skewed, while the BPVS controls showed a bi-modal distribution).

n_1=15, n_2=15, p=0.05 (Bonferroni corrected α=0.008). Again, the TROG and BPVS controls did not differ significantly from each other, W=194.5, n_1=15, n_2=15, p=0.11.

Thus, in summary, the Age controls had higher BPVS and TROG raw scores than the participants with SLI and the language controls. The two language control (TROG and BPVS) groups did not differ significantly from each other on the language tests or age. The relationship between the SLI group and the BPVS versus TROG controls was also the same. For these reasons, the TROG and BPVS control groups were merged to form a larger 'language control group'. This group consisted of 30 children, fifteen matched to the SLI group on the TROG and fifteen on the BPVS.

Non-standardized assessments

Non-word repetition: Test of Phonological Structure (TOPhS, van der Lely & Harris, 1999).

The TOPhS consists of ninety-six non-words which are derived from four basic non-words /ˈdepə/, /ˈpɪfi/, /ˈketə/, /ˈfɪpə/ with the simplest stress pattern (Strong Weak – SW) and no consonant clusters, thus fitting the template CVCV. The 24 non-words based on the basic non-word /ˈdepə/ are shown in Table 2. The structure of the NWs based on the other three basic words vary in the same ways as those based on /ˈdepə/. The variants of the basic non-words were constructed using five binary phonological parameters, three of which control aspects of 'syllabic structure' (including the addition of consonants to form consonant clusters such as 'pr' and 'kl') and two of which control aspects of 'metrical structure' (involving stress patterns). Each parameter has two options: marked or unmarked (see Harris, 1994). The unmarked structure occurs in all languages and appears early in the phonological acquisition process whereas the marked structure occurs only in a subset of languages and is acquired relatively late[3].

Of the three parameters controlling syllable structure, one establishes whether an ***onset*** contains one consonant (unmarked, e.g., /ˈdepə/) or a consonant cluster (marked, e.g., /ˈdrepə/). Another determines whether a ***rhyme*** is open, i.e. ends in a vowel (unmarked e.g., /ˈdepə/) or is closed, i.e. ends in a consonant (marked e.g., /ˈdempə/). A third establishes whether a ***word end***s in a vowel (unmarked, e.g., /ˈdepə/) or a consonant (marked, e.g., /ˈdep/). The two parameters controlling metrical structure establish whether a weak syllable is adjoined to the beginning (***Left Adjunction***) or the end of the NW (***Right Adjunction***). Adjunction adds to the metrical complexity of a word and constitutes the marked option. Left Adjunction leads to the marked stress patterns WS: (e.g., /bəˈdep/) or WSW: (e.g., /bəˈdepə/) and Right Adjunction to the marked pattern SWW (e.g., /ˈdepəri/). Words with both Left and Right Adjunction are the most complex metrically and have the marked stress pattern WSWW (e.g., /bəˈdepəri/).

The five parameters of the TOPhS allow us to look at the effects of phonological complexity in general (by counting how many of these five parameters are 'marked'), or specifically at 'syllabic complexity' (the number of marked syllable parameters, Onset, Rhyme and Word End) or 'metrical complexity' (the number of marked metrical parameters, Left and Right Adjunction). To analyse the effect of length, we counted the number of syllables in each NW. Table 2 shows (for non-words based on /ˈdepə/) which parameters are marked (hi-lighted in bold and underlined in non-word) and also gives a count for overall phonological complexity, syllabic complexity and metrical complexity as well as the number of syllables.

TABLE 2 ABOUT HERE

[3] For more details see Gallon et al. (2007), and for further theoretical background see Harris (1994).

The relationship between length and complexity on the TOPhS is shown in Table 3. This shows the number of NWs which have a particular number of marked structures and number of syllables. The correlation between the number of marked structures (phonological complexity) and the number of syllables (length) is r=.27, p=.009.

TABLE 3 ABOUT HERE

The participants were told they would hear some “funny, made up words”, which they should try to repeat into a microphone. The list of TOPhS non-words was audio recorded by a female native English speaker using a Bruel and Kjaer sound level meter (model 2231). A 3 second silent interval occured after each non-word. The children heard the digitally recorded NWs via Sennheiser AD475 headphones and their repetitions were recorded onto a DAT tape (TCD-D8) via an external Sony Electret condenser microphone. The NWs were presented without breaks in a set random order. Repetitions were transcribed on-line and subsequently verified against the recording by the first author. We primarily used the total number of words correct (TNC), rather than a percent phoneme correct (PPC) score, as we were mainly interested in the overall phonological structure of the word. Furthermore, when Graf Estes et al. (2007) compared TNC versus PPC scoring for the NRT they found TNC had a lower effect size and thus was more conservative. However, we also re-analysed our results with PPC scoring to see if this had any influence on the results. Inter-rater reliability was computed by comparing the transcriptions for four randomly selected children with SLI (26%) with those of an independent transcriber working from the digital recording. Children with SLI were used for reliability testing as they made more errors than controls and were also deemed more difficult to transcribe. Phoneme by phoneme inter-rater agreement for these four transcriptions was 99%.

Verb Agreement and Tense Test (VATT, van der Lely, 2000)

The VATT considers a child's ability to use past tense and 3rd person singular agreement for ten regular and ten irregular verbs (each group containing five high and five low frequency verbs), and balanced for neighbourhood effects. The participants completed forty sentences to describe twenty picture stimuli in both the present tense, requiring 3rd person singular agreement (e.g., *every day Woody slams the door*) and past tense (e.g., *yesterday Woody slammed the door*). In this study, we scored past tense responses as correct if tense was marked, even if this meant over-regularisation of irregular verbs (e.g., *maked*, *gived).* Thus, we assessed syntactic tense marking per se, rather than the lexical-morphological stored knowledge of irregular past tense forms. The addition of 3s and past tense morphemes can lead to word final clusters which are phonologically more complex. In the VATT, a cluster is required for correct production of agreement[4] for 17/20 verbs, and of past tense for 11/20 verbs. Therefore the phonological demands are greater for agreement than past tense production in this test.

Results

Tests of normality (Shapiro-Wilk) showed that scores on the TOPhS for both the age controls and the SLI group deviated significantly from normality (Age: p=.003, SLI: p=.03). In the case of the age controls this was due to one outlier who scored significantly lower than the others. When this score was removed, the age control data were normally distributed (p=.30). The presence or absence of this score made no difference to the significance of any of the results, so this participant's scores were included in all analyses. In the case of the children with SLI, a histogram (Figure 1) shows the deviation from normality was due to a bi-modal distribution where no participant scored between 47 and 68 correct out of 96.

FIGURE 1 ABOUT HERE

[4] We note that in English the "s" morpheme conflates both tense and agreement marking.

The children with SLI therefore cannot be considered to be a homogeneous group with respect to their NWR abilities. For this reason we split them into two groups on the basis of their scores on the TOPhS: 'SLI-high' (8 participants, mean: 79, SD: 9.3) and 'SLI-low' (7 participants, mean: 39, SD: 6.2). Shapiro-Wilk tests on these two groups were non-significant indicating normal distributions. These smaller groups were used for all further analyses. The control participants had been individually matched to the participants with SLI, so it was possible to create new control groups for the two new SLI groups. The controls who had been individually matched to the participants in the SLI-low group on language formed a new Language Control-low group (henceforth 'LC-low') and those who had been matched on age formed a new Age Control-low group (henceforth 'AC-low'). Those controls individually matched to the participants in the SLI-high group formed new Language Control-high (henceforth 'LC-high') and Age Control-high (henceforth AC-high) groups.

A boxplot (Figure 2) of TOPhS scores for the two SLI groups and their individually matched language and age controls shows the SLI-low group achieved much lower scores on the TOPhS than any of the other groups. Indeed there was no overlap between the SLI-low group and the other five groups. In contrast, the SLI-high group achieved similar scores to all control groups.

FIGURE 2 ABOUT HERE

The TOPhS scores of each of the two SLI groups were compared with those of their individually matched controls using one-way ANOVAs with Group as the between subjects variable (SLI, LC, AC). When comparing the 'SLI-high' group with their individual matches, no significant effect of Group was found, $F(2,29)=0.64$, $p=0.53$, $\eta p^2=0.04$. In contrast, a highly significant effect of Group was found for the 'SLI-low' group and their individually matched controls, $F(2,25)=151.13$, $p<0.001$, $\eta p^2=0.92$. Post-hoc Bonferroni corrected t-tests showed that this was due to highly significant differences with very large effect sizes between the 'SLI-low' group and both control groups (SLI-low vs. LC-low: $p<0.001$, $d=7.08$; SLI-low vs. AC-low: $p<0.001$, $d=9.64$). These findings are particularly striking when we consider the ages of the children: the SLI-low group ranged in age from 11;4 to 14;8 and all scored well below the youngest control (age: 5;4, score: 66) despite being 6 to 9 years older. Exactly the same pattern of similarities and differences were found using PPC scoring, the effect sizes for the comparisons of the SLI-low group using PPC scoring were SLI-low vs. LC-low: $d=4.22$; SLI-low vs. AC-low: $d=4.45$.

Comparison of the two SLI groups

To investigate further the nature of the differences between the two SLI groups, we compared their ages and scores on the TOPhS, VATT and the standardized language, literacy and performance tests using 2-tailed t-tests (or Wilcoxon signed ranks tests for non-normally distributed data). Where raw scores were inappropriate, z-scores were used. Table 4 shows these data along with the *p*-values and effect sizes (*d*) for t-tests. The two groups differed significantly (with large effect sizes) on the TOPhS and VATT (total, agreement and tense), but not on age or any of the other standardized tests. Despite no significant differences between the two SLI groups on the WORD, all three with normal scores on Reading accuracy were in the SLI-high group. Of the additional two who had normal scores on Spelling accuracy, but not Reading accuracy, one was in the SLI-low and one in the SLI-high group.

TABLE 4 ABOUT HERE

As the VATT is not standardized, it is important to establish whether either or both of the two groups of participants with SLI in fact had difficulties on this measure and whether this impairment was greater than would be expected given their general language levels. Initially data were collected for the participants with SLI and their language controls only. The SLI-low group showed no overlap with their language controls, whereas the SLI-high group showed a very high degree of overlap with their language controls (see Table 5). Therefore data were also collected for the age controls for the SLI-high group. It was not necessary to collect data for the age controls for the SLI-low group as they were obviously more impaired that their language controls, so a difference from their age controls was assumed. Non-parametric tests were used to compare the SLI groups with their control groups and the resulting *p*-values are shown in Table 5.

TABLE 5 ABOUT HERE

The results show that the SLI-low group performed much worse than their language controls on both tense and agreement. In contrast, the SLI-high group had significantly poorer performance on tense but not agreement than their language controls and on both tense and agreement than their age controls. We also analysed whether within each group, performance differed between tense and agreement. The *p*-values in Table 5 showed that the SLI-low group were significantly worse at agreement than tense, while their language controls showed the opposite pattern. The SLI-high group and their controls showed no significant differences between tense and agreement.

At the recruitment stage we had analysed whether the children with SLI differed from their controls on the BPVS and TROG. However, this was done with the whole SLI group, not the smaller SLI-low and SLI-high groups. These smaller groups should not differ from their language controls since they were individually matched on these tests, but we wanted to ensure that both SLI groups (as well as the overall combined group) were indeed impaired relative to age controls on the BPVS and TROG. *T*-tests confirmed that both SLI groups differed from their age controls, scoring significantly lower on both the BPVS (SLI-low: $p<0.001$; SLI-high: $p=0.007$) and TROG (SLI-low: $p=0.001$; SLI-high: $p=0.03$).

In summary, the SLI-low group were impaired compared to age controls on both the BPVS and TROG tests, while having equal scores on these tests to their language controls.

However, they were worse than their language controls on the VATT and the TOPhS tests. The SLI-high group scored worse than language controls on tense, but not agreement, BPVS, TROG or TOPhS. However, they were impaired relative to age controls on the VATT, BPVS, and TROG, but not the TOPhS, despite scoring at least 1.5 standard deviations below the mean on both receptive and expressive language measures on the CELF-3. The two SLI groups did not differ significantly from each other on the BPVS, TROG, CELF-3 or WORD; only on the VATT and the TOPhS.

Relationship between TOPhS and other language and literacy measures

In order to investigate the relationship between performance on the TOPhS and other language and literacy measures in both typical development and in SLI, separate correlations were performed for the 45 TD controls and 15 children with SLI. Spearman rank correlations were used for the non-normal SLI data, while Pearson r and partial correlations (partialling out age which was significantly correlated with everything) were used for the normally distributed TD control data. The results are shown in Table 6 and a scatterplot showing the relationship between the TOPhS and the VATT for the two groups is in Figure 3.

TABLE 6 ABOUT HERE

FIGURE 3 ABOUT HERE

For TD controls, only the partial correlations with TOPhS PPC and the BPVS were significant (see Table 6). For the children with SLI, their TOPhS TNC scores did not correlate with age and correlations with the standardised language and literacy measures did not reach significance. However, their TOPhS TNC scores showed very strong, highly significant correlations with PPC scores and with the VATT (total, agreement and tense) scores. Indeed the correlation with agreement was almost perfect (r=.97). This contrasts with much lower nonsignificant correlations between the TOPhS and VATT in the controls once age had been partialled out (*r*=.16 for agreement). This lack of significant correlations could be due to ceiling effects in the controls on the VATT (see Figure 3). Therefore the TD correlations were repeated, excluding all data from participants who scored 39 or 40 (out of 40) on the VATT. The TNC scores on the TOPhS of those excluded ranged from 76-96 correct (out of 96). This left only 22 TD participants, whose TOPhS TNC scores ranged from 61-95 correct. The Pearson correlations for the remaining TD participants were not significant (Total: r=.33, p=.13, tense: r=.36, p=.11, agreement: r=.23, p=.30), and neither were the partial correlations, partialling out the effect of age (Total: r=.19, p=.41, tense: r=.21, p=.36, agreement: r= .11, p=.62).

Influence of length and phonological complexity

Next we compared the TD group (n=45) and the SLI-low (n=7) and SLI-high (n=8) groups on the factors inherent in the non-words which may influence performance on the TOPhS. We considered the relationship between performance and two factors underlying the non-words: length and phonological complexity. All analyses in this section were carried out by-items, where the dependent variable was the proportion of participants in each group correctly repeating each non-word (whole word score). We also looked at the mean PPC for each non-word.

FIGURES 4 AND 5 ABOUT HERE

The graphs in Figures 4 and 5 show the influence of length (number of syllables, Figure 4) and phonological complexity (number of marked structures, Figure 5) on the TOPhS whole word performance of the three groups. The similarity of the two graphs is striking. Performance appears to be related to both factors, with the SLI-low group generally scoring lower, but also affected more by increasing length and complexity than the SLI-high and control groups, as shown by the steeper slopes of the lines. We analysed the data using Spearman (1-tailed)[5] rank correlations (due to the strong ceiling and floor effects). Table 7 shows the rank correlations (*r*)

[5] We used one-tailed tests because the previous literature indicates that any relationship between the two measures is negative, thus we were only interested in a relationship in this direction.

for each of the three groups of participants on each variable. Length was measured in the number of syllables and phonological complexity in the number of marked structures. Phonological complexity was also divided into metrical versus syllabic complexity.

TABLE 7 ABOUT HERE

The results in Table 7 show significant relationships for all three participant groups between the TOPhS whole word scores and both length and phonological complexity. However, when phonological complexity was divided into metrical versus syllabic complexity, the relationship with syllable complexity was not significant for any group. However, syllable complexity on the TOPhS consists of three parameters: Onset, Rhyme and Word End. The effect of Word End has not been studied previously. Therefore we created a new variable: 'Added consonants', the number of additional consonants added (Onset and Rhyme). This was negatively correlated with performance in both SLI groups. In contrast, Word End showed a positive correlation for the SLI groups. Therefore the non-significant correlation of syllable complexity with performance of the participants with SLI was due to the positive correlation of the Word End parameter in conjunction with the negative correlation of the Onset and Rhyme parameters. The negative correlation of performance with the new Added Consonants variable is also in line with previous research.

The PPC scoring showed similar patterns for the SLI groups, where Added consonants, metrical complexity and length (as measured by number of syllables) were significantly related to TOPhS performance. However, using PPC scoring, the controls showed no significant relationships between TOPhS performance and any of the factors analysed.

Thus, in summary, using whole word scoring (proportion of participants correctly repeating each NW) the performance of all groups was negatively correlated with both length and metrical complexity and the performances of the SLI groups were also negatively correlated with the number of consonants added to the stressed syllable (Added consonants: Onset and Rhyme).

These factors predicted performance to the greatest extent in the SLI-low group and least in the control group. Using PPC scoring, the SLI groups showed a similar pattern as with whole word scoring, but the controls showed no obvious pattern.

Because metrical complexity and number of syllables are strongly correlated (*r*=.89, *p*<.001) in the TOPhS, it is not clear whether either factor has an independent effect on performance of all groups using whole word scoring or the SLI groups using PPC. Therefore, the final analyses considered the independent effects of length, metrical complexity and 'Added consonants' (Onset and Rhyme) for each group by carrying out one-tailed partial correlations for each of these with the TOPhS, partialling out the effects of the other two factors[6]. The results are also shown in Table 7. These showed that length was not independently correlated with performance for any group, using either scoring system; metrical complexity and Added consonants were independently correlated with performance for both SLI groups using both scoring systems; for the controls, metrical complexity was negatively correlated with whole word TOPhS performance but there were no significant correlations using PPC scoring.

Discussion

The participants with SLI in this study showed a bi-modal distribution in their ability to repeat the non-words of the TOPhS. Half achieved a similar degree of accuracy to TD control groups while the other half scored significantly worse. The effect size for the difference between our SLI-low group and their age controls using TNC scoring (*d*=9.64) was more than double the largest effect size (*d*=4.34) reported in the Graf-Estes et al. (2007) meta-analysis, although using PPC scoring we found an effect size of *d*=4.45. Our finding of smaller effect sizes with PPC scoring is the opposite pattern to that in Graf-Estes et al. (2007). However, they used the NRT which differs greatly from the TOPhS in construction, particularly in that it does not include any

[6] It was not possible to carry out multiple regression because several assumptions necessary for reliable interpretation of the results were violated, including normality in arrays and multiple collinearity

consonant clusters. Such a clear division of performance as we found within a group of children diagnosed with SLI has not been shown in previous studies. However, the division into children with high versus low performance on the TOPhS is probably responsible for the very large effect sizes for the differences between the SLI-low group and all other groups.

Table 8 summarizes our findings. The two SLI groups (low vs. high TOPhS) did not differ significantly on any other language or literacy measures except verb agreement and tense marking (VATT). Comparisons with age controls showed that while the SLI-high group had no detectable impairment in repeating non-words (TOPhS), both SLI groups were impaired in production of verb agreement and tense marking (VATT) and understanding of vocabulary (BPVS) and sentences (TROG). Z-scores on the CELF-3 showed that all the participants with SLI were impaired on this general receptive and expressive language test and their z-scores on the WORD showed that on average they also were impaired on literacy measures. However, three participants with SLI scored within the normal range on Reading Accuracy (who were all in the SLI-high group) and five on Spelling Accuracy (one of whom was in the SLI-low group).

TABLE 8 ABOUT HERE

Comparisons of the SLI participants with their language controls revealed particular difficulties over and above their general language difficulties for the SLI-low group with verb agreement and tense marking (VATT) and repeating non-words (TOPhS) and for the SLI-high group with tense marking only.

Correlations within the 15 children with SLI (SLI-low and SLI-high groups combined) versus the 45 TD controls (all control groups combined) showed the TOPhS TNC scoring correlated significantly with 1) the TOPhS PPC scoring for both groups, 2) the BPVS only for the controls and 3) the VATT only for the participants with SLI. The low number of participants with SLI may be responsible for the lack of significance of some of their correlations. However, despite their low numbers, the correlation between their TOPhS and VATT scores was still highly significant.

Good NWR performance among children with SLI

Only half of the participants with SLI were impaired on the TOPhS (the SLI-low group) despite all the participants with SLI having low language scores on the CELF-3. The finding that some children with SLI do well on the TOPhS concurs with previous studies showing that some children with SLI pass NWR tests (Bishop et al., 2006; Botting & Conti-Ramsden, 2001; Chiat & Roy, 2007; Gardner et al., 2006). The practical implication is therefore that NWR tasks should not be used in isolation as a screen for identifying SLI.

It has been proposed that the presence or absence of co-occurring dyslexia could account for variation in performance in children with SLI (Catts et al., 2005). Indeed, studies have found that children with SLI only without dyslexia, did not differ significantly from age controls (Bishop et al., 2009; Rispens & Parigger, 2010). Those studies split the participants with SLI by reading abilities and found significant differences in NWR abilities. In contrast, we split our participants by NWR performance and found no significant differences in their reading abilities. Three of our SLI-high group had reading abilities within the normal range, but the other five had impaired reading accuracy (with good NWR). Therefore, our findings suggest that while children with SLI without dyslexia probably do have good NWR abilities, as suggested by Catts et al. (2005), the reverse is probably untrue (i.e., that good NWR abilities predict good reading abilities) because some of our SLI-high group had poor reading abilities despite scoring well on NWR.

Given the older age of our participants, it is possible that the SLI-high group had had difficulties with phonology and/or other factors underlying NWR in the past which had resolved while their impairment in other language areas persisted. However, this interpretation is inconsistent with the previous findings that while some children with SLI appear to outgrow their general language difficulties, they remain impaired in NWR (Bishop et al., 1996) suggesting this is a persistent deficit. Thus, it seems unlikely that the current language difficulties of our SLI-high group are due to difficulties with phonology, phonological short-term memory or storage, or lower level processes involved in speech perception and processing, as all of these would also be expected to impair their ability to repeat non-words. Thus, some other factor must underlie their language difficulties.

Some theories of SLI provide possible non-phonological explanations, particularly for difficulties with verb agreement and tense marking: these include difficulties establishing agreement (Clahsen, 1989) or knowing that tense and agreement are obligatory in matrix clauses (Rice et al., 1995). The impairment on tense marking (even relative to language controls) of all the children with SLI supports Rice and Wexler's (1996) hypothesis of an extended optional infinitives stage in SLI. However, given that our participants with SLI were in their teenage years, this problem does not appear to be one of delay, but to be part of an ongoing developmental impairment and fit with Rice, Hoffman and Wexler's (2009) recent findings of a lower asymptote in children with SLI. However, these theories do not account for the reduced understanding of vocabulary and sentences of the SLI participants. One hypothesis which can explain difficulties understanding sentences, is van der Lely's (2005) Computational Grammatical Complexity (CGC) hypothesis, which was developed from her Representational Deficit for Dependent Relations hypothesis (RDDR, van der Lely, 1998). The CGC hypothesis puts forward that abstract hierarchical structures are impaired in children with SLI and that different components of grammar may be independently impaired but have a cumulative impact on performance. Thus, the SLI-high group could be assumed to have difficulties with syntax and morphology, but have spared phonology, with vocabulary difficulties arising from difficulties using syntax to aid vocabulary learning (van der Lely, 1994).

Relationship between NWR and other language and literacy measures

The SLI-low and SLI-high groups differed significantly on only two tests: the TOPhS and the VATT (both tense and agreement). Within the whole SLI group (SLI-low and SLI-high groups combined), significant correlations were also found between these measures. However, among the TD controls, the correlations in our study were weak and non-significant and appeared to be largely mediated by age. The lack of correlations could have been due to a ceiling effect in the controls on the VATT test, but this persisted when those scoring near ceiling were excluded. Thus, performance on the TOPhS appears to be related to syntactic and morphological agreement and tense marking in sentences in children with SLI, but not in TD children.

Correct production of many items on the VATT requires production of a word final consonant cluster (e.g., *drops, dropped*), more for agreement than tense. Marshall and van der Lely (2007) showed children with SLI (but not TD children) are less likely to suffix stems when the inflected form ends in a cluster. We showed in this study that the addition of consonants to form clusters affected the performance of both our SLI groups (but not the TD children) on the TOPhS, with the SLI-low group affected to the greatest extent. Therefore, a significant correlation between the TOPhS and VATT tests amongst the children with SLI, but not TD children is unsurprising. Indeed, the SLI-low group were the only group who had worse performance on agreement (which required more clusters) than tense.

We therefore propose that the SLI-low group have an additional deficit in phonology which impacts on verb tense and agreement marking, especially where clusters are required, resulting in poorer performance in this group on the VATT and a strong relationship between the TOPhS and the VATT for the whole SLI group.

We found a weaker, but significant correlation between NWR and vocabulary in the TD controls. The correlations for the SLI group on vocabulary, TROG and literacy measures failed to reach significance. This could be in part due to low numbers of participants, but our findings are in line with many previous studies which found correlations between vocabulary and NWR in TD children (e.g., Briscoe et al., 2001; Gathercole & Baddeley, 1989) but not children with SLI (e.g., Briscoe et al., 2001; Edwards & Lahey, 1998; Gallon et al., 2007). Other studies also found a weak, but non-significant correlation with the TROG (e.g., Gallon et al., 2007; Montgomery, 2004) and no significant correlations between literacy measures and NWR in children with SLI (de Bree et al., 2010).

Influence of length and phonological complexity

When we considered the effects of non-word length and phonological complexity, we found a strong correlation between NWR performance on the TOPhS and both length and phonological complexity for the participants with SLI; thus, these data could support all current theories of the underlying reasons for difficulty with NWR tasks among children with SLI: poor phonological storage (Gathercole, 2006), or difficulties forming phonological representations (Chiat, 2001; Edwards & Lahey, 1998). However, length and complexity are related, so we used partial correlations to establish the independent contribution to the variance of each measure. These revealed that the performance of the SLI participants was related to both metrical complexity and the presence of consonant clusters. The effect of metrical complexity supports Chiat's (2001) view that some children with SLI have particular difficulties identifying the phonological segmental details in rhythmic structures. The effect of consonant clusters supports the view that at least some children with SLI have difficulties with phonologically complex representations (Gallon et al., 2007; Marshall & van der Lely, 2007; 2009).

When the effects of consonant clusters and metrical complexity were partialled out, length (in terms of the number of syllables) was no longer significantly related to performance. Thus, our results suggest that the TOPhS test and possibly also other tests of NWR should not be primarily viewed as tests of short-term memory or phonological storage as shown by a length effect. If our findings from the partial correlations can be replicated in larger studies specifically designed to contrast the influence of number of syllables versus phonological complexity, this would indicate that SLI is less likely to be caused by limited phonological short-term memory capacity or impaired phonological storage.

Summary

The number of participants in this study is small. This may mean that we were unable to detect some smaller differences between groups. However, our numbers were sufficient to detect large and highly significant differences between groups on some measures. If our findings can be replicated in larger studies, they raise important questions for theories of SLI. Although theories proposing a single underlying deficit for SLI may be more parsimonious, those currently available cannot account for the data in this study. Several researchers have proposed that dissociable deficits or risk factors may better account for the data. Bishop et al. (1999) proposed that a series of risk factors may be involved in SLI and those children with more than one risk factor are likely to show a more severe deficit. Bishop and Snowling (2004) proposed that children with classic dyslexia have poor phonological skills; poor reading comprehenders have poor semantic skills; but children with SLI have both poor phonological and semantic skills. Van der Lely (2005) proposed the Computational Grammatical Complexity Hypothesis (CGC) whereby children with SLI may have independent but interactive deficits in any one or a combination of phonology, morphology and/or syntax. In order to account for our data, we propose that all participants with SLI have impairments in syntax and morphology which affect their general language abilities, including marking of agreement and tense. However, we propose that the SLI-low group have an additional phonological impairment (which the SLI-high group do not have) which greatly affects their ability to repeat the non-words on the TOPhS and further impairs their ability to mark tense and particularly agreement in sentences accurately.

Conclusions

Our main finding is that the participants with SLI fell into two distinct groups. They all had significant syntactic and morphological difficulties with marking agreement and tense and also had severe difficulties with a general language test (the CELF-3) but only half had difficulties repeating the non-words of the TOPhS. If our findings can be replicated in larger studies, this would indicate the TOPhS and probably NWR tests in general should not be used in isolation to identify language impairments. Also, those factors which have been proposed to account for NWR difficulties (e.g., difficulties with phonology, phonological short-term memory or storage, or with processing phonology, speech or brief acoustic cues) cannot account for the language difficulties of half of our participants who had no detectable difficulties with NWR (the SLI-high group), although they may account for some of the difficulties with verb tense and particularly agreement marking of the other half (the SLI-low group). Our preliminary finding that phonological complexity (but not number of syllables) had an independent effect on performance suggests that the difficulties some children with SLI have with NWR are more likely to be due to difficulties with phonology per se rather than phonological short-term memory or phonological storage.

Acknowledgements

This research was supported by Moor House School, Surrey, UK and Wellcome Trust Grant 063713. We would like to thank Chloe Marshall for carrying out the interrater reliability and carrying out some of the language tests on some control children and Chloe Marshall and Nichola Gallon for many useful discussions. We would like to thank the children who took part in the study and their schools: Moor House School, Ifield First and Middle Schools, Southgate West First and Middle Schools, Brockham Primary School and Oxted Secondary School. Heather van der Lely declares a personal interest in the VATT test. No other author associated with

research involving these tests has any conflict of interest.

Reference List

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

Baird, G., Slonims, V., Simonoff, E., & Dworzynski, K. (2011). Impairment in non-word repetition: a marker for language impairment or reading impairment? *Developmental Medicine and Child Neurology, 53,* 711-716.

Bishop, D. V. M. (1979). Comprehension in developmental language disorders. *Developmental Medicine and Child Neurology, 21,* 225-238.

Bishop, D. V. M. (1989). Test of Reception of Grammar. Oxford, UK, Available from the author.

Bishop, D. V. M., Adams, C. V., & Norbury, C. F. (2006). Distinct genetic influences on grammar and phonological short-term memory deficits: evidence from 6-year-old twins. *Genes, Brain and Behavior, 5,* 158-169.

Bishop, D. V. M., Carlyon, R. P., Deeks, J. M., & Bishop, S. J. (1999). Auditory temporal processing impairment: Neither necessary nor sufficient for causing language impairment in children. *Journal of Speech Language and Hearing Research, 42,* 1295-1310.

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

Bishop, D. V. M. & Snowling, M. (2004). Developmental dyslexia and Specific Language Impairment: same or different? *psychological bulletin, 130,* 858-886.

Bishop, D. V. M., McDonald, D., Bird, S., & Hayiou-Thomas, M. E. (2009). Children Who Read Words Accurately Despite Language Impairment: Who Are They and How Do They Do It? *Child Development, 80,* 593-605.

Botting, N. & Conti-Ramsden, G. (2001). Non-word repetition and language development in children with specific language impairment (SLI). *International Journal of Language & Communication Disorders, 36,* 421-432.

Briscoe, J., Bishop, D. V. M., & Norbury, C. F. (2001). Phonological processing, language, and literacy: A comparison of children with mild-to-moderate sensorineural hearing loss and those with specific language impairment. *Journal of Child Psychology and Psychiatry and Allied Disciplines, 42,* 329-340.

Catts, H. W., Adlof, S. M., Hogan, T. P., & Weismer, S. E. (2005). Are specific language impairment and dyslexia distinct disorders? *Journal of Speech Language and Hearing Research, 48*, 1378-1396.

Chiat, S. (2001). Mapping theories of developmental language impairment: Premises, predictions and evidence. *Language and Cognitive Processes, 16,* 113-142.

Chiat, S. & 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.

Clahsen, H. (1989). The Grammatical Characterization of Developmental Dysphasia. *Linguistics, 27,* 897-920.

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

de Bree, E., Wijnen, F., & Gerrits, E. (2010). Non-word Repetition and Literacy in Dutch Children At-risk of Dyslexia and Children with SLI: Results of the Follow-up Study. *Dyslexia, 16*, 36-44.

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, L. M., Whetton, C., & Burley, J. (1997). The British Picture Vocabulary Scale, Second Edition. Windsor, NFER-NELSON.

Edwards, J. & Lahey, M. (1998). Nonword repetitions of children with specific language impairment: Exploration of some explanations for their inaccuracies. *Applied Psycholinguistics, 19*, 279-309.

Elliot, C. D., Smith, P., & McCulloch, K. (1996). *British Ability Scales II*. Windsor: NFER-NELSON.

Farmer, M. (2000). Language and social cognition in children with specific language impairment. *Journal of Child Psychology and Psychiatry and Allied Disciplines, 41,* 627-636.

Gallon, N., van der Lely, H. K. J., & Harris, J. (2007). Non-word repetition: an investigation of phonological complexity in children with Grammatical SLI. *Clinical Linguistics & Phonetics, 21,* 435-455.

Gardner, H., Froud, K., McClelland, A., & van der Lely, H. K. J. (2006). Development of the Grammar and Phonology Screening (GAPS) test to assess key markers of specific language and literacy difficulties in young children. *International Journal of Language & Communication Disorders, 41,* 513-540.

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

Gathercole, S. E. & Baddeley, A. D. (1989). Evaluation of the role of phonological STM in the development of vocabulary in children: a longtitudinal study. *Journal of Memory and Language, 28,* 200-213.

Gathercole, S. E. & Baddeley, A. D. (1990). Phonological memory deficits in language disordered children: is there a causal connection? *Journal of Memory and Language, 29,* 336-360.

Graf Estes, K., Evans, J. L., & Else-Quest, N. M. (2007). Differences in the nonword repetition performance of children with and without specific language impairment: a meta-analysis. *J Speech Lang Hear Res, 50,* 177-195.

Hansson, K., Forsberg, J., Lofqvist, A., Maki-Torkko, E., & Sahlen, B. (2004). Working memory and novel word learning in children with hearing impairment and children with specific language impairment. *International Journal of Language & Communication Disorders, 39,* 401-422.

Harris, J. (1994). *English Sound Structure*. Oxford: Blackwell.

Leonard, L. B., Bortolini, U., Caselli, M. C., McGregor, K. K., & Sabbadini, L. (1992). Morphological deficits in children with specific language impairment: The status of features in the underlying grammar. *Language Acquisition, 2,* 151-179.

Marshall, C. R. & van der Lely, H. K. J. (2007). The impact of phonological complexity on past tense inflection in children with Grammatical-SLI. *Advances in Speech Language Pathology, 9,* 191-203.

Marshall, C. R. & 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.

Montgomery, J. (2004). Sentence comprehension in children with specific language impairment: effects of input rate and phonological working memory. *International Journal of Language & Communication Disorders, 39,* 115-133.

Montgomery, J. W. (1995). Sentence Comprehension in Children with Specific Language Impairment - the Role of Phonological Working-Memory. *Journal of Speech and Hearing Research, 38,* 187-199.

Montgomery, J. W. & Windsor, J. (2007). Examining the language performances of children with and without specific language impairment: Contributions of phonological short-term memory and speed of processing. *Journal of Speech Language and Hearing Research, 50,* 778-797.

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

Norbury, C. F., Bishop, D. V. M., & Briscoe, J. (2001). Production of English finite verb morphology: A comparison of SLI and mild-moderate hearing impairment. *Journal of Speech Language and Hearing Research, 44,* 165-178.

Rice, M. L. & Wexler, K. (1996). Toward tense as a clinical marker of specific language impairment in English-speaking children. *Journal of Speech and Hearing Research, 39,* 1239-1257.

Rice, M. L., Wexler, K., & Cleave, P. L. (1995). Specific Language Impairment As A Period of Extended Optional Infinitive. *Journal of Speech and Hearing Research, 38,* 850-863.

Rice, M. L., Hoffman, L., & Wexler, K. (2009). Judgments of Omitted BE and DO in Questions as Extended Finiteness Clinical Markers of Specific Language Impairment (SLI) to 15 Years: A Study of Growth and Asymptote. *Journal of Speech Language and Hearing Research, 52*, 1417-1433.

Rispens, J. & Parigger, E. (2010). Non-word repetition in Dutch-speaking children with specific language impairment with and without reading problems. *British Journal of Developmental Psychology, 28,* 177-188.

Rust, J., Golombok, S., & Trickey, G. (1993). Wechsler Objective Reading Dimensions. Sidcup, UK, Psychological Corporation.

Sahlen, B., Reuterskiold-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,* 337-352.

Semel, E., Wiig, E. H., & Secord, W. A. (1995). *Clinical Evaluation of Language Fundamentals-3*. San Antonio: The Psychological Corporation.

Snowling, M., Chiat, S., & Hulme, C. (1991). Words, nonwords, and phonological processes: Some comments on Gathercole, Willis, Emslie, and Baddeley. *Applied Psycholinguistics, 12,* 369-373.

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.

van der Lely, H. K. J. (1994). Canonical Linking Rules - Forward Versus Reverse Linking in Normally Developing and Specifically Language-Impaired Children. *Cognition, 51,* 29-72.

van der Lely, H. K. J. (1998). SLI in Children: Movement, Economy, and Deficits in the Computational-Syntactic System. *Language Acquisition, 7,* 161-192.

van der Lely, H. K. J. (2000). Verb Agreement and Tense Test (VATT).   available from www.dldcn.com.

van der Lely, H. K. J. (2005). Domain-specific cognitive systems: Insight from Grammatical-specific language impairment. *Trends in Cognitive Sciences, 9,* 53-59.

van der Lely, H. K. J. & Harris, J. (1999). Test of Phonological Structure. Available from www.dldcn.com.

Table 1: *Mean (SD) plus ranges on matching criteria (raw scores for BPVS and TROG, z-scores for BAS, years;months for Age).*

	SLI	**TROG controls**	**BPVS controls**	**Age controls**
TROG	15.40 *(2.32)* 9 to 18	15.40 *(2.32)* 9 to 18	17.00 *(1.69)* 15 to 19	18.33 *(0.90)* 17 to 20
BPVS	91.07 *(14.24)* 63 to 115	87.00 *(16.17)* 58 to 120	90.87 *(13.84)* 65 to 112	121.87 *(13.45)* 99 to 149
BAS	-0.04 *(0.82)* -0.95 to1.55	0.33 *(0.60)* -0.60 to 1.25	0.53 *(0.68)* -0.55 to 1.95	0.06 *(0.66)* -1.15 to 1.60
Age	13;1 *(1;3)* 11;0 to 14;11	8;3 *(1;8)* 5;4 to 11;3	8;8 *(1;6)* 5;10 to 12;2	13;1 *(1;3)* 11;3 to 14;10

Table 2: *Structure of the TOPhS test (van der Lely & Harris, 1999) using one of the four basic words as an example. Three other basic words were used in the same way to create the remaining non-words; u=unmarked, m=marked*

						Syllable parameters				**Metrical parameters**			Total marked structures	Number of syllables
TOPhS non-word						Onset	Rhyme	Word End	Number marked	Left Adj	Right Adj	Number marked		
	d	e	p	e		u	u	u	0	u	u	0	0	2
	dr	e	p	e		m	u	u	1	u	u	0	1	2
	d	e	**mp**	e		u	m	u	1	u	u	0	1	2
	d	e	p	_		u	u	m	1	u	u	0	1	1
be	d	e	p	e		u	u	u	0	m	u	1	1	3
	d	e	p	e	**ri**	u	u	u	0	u	m	1	1	3
	dr	e	**mp**	e		m	m	u	2	u	u	0	2	2
	dr	e	p	_		m	u	m	2	u	u	0	2	1
	d	e	**mp**	_		u	m	m	2	u	u	0	2	1
be	**dr**	e	p	e		m	u	u	1	m	u	1	2	3
	dr	e	p	e	**ri**	m	u	u	1	u	m	1	2	3
be	d	e	**mp**	e		u	m	u	1	m	u	1	2	3
	d	e	**mp**	e	**ri**	u	m	u	1	u	m	1	2	3
be	d	e	p	_		u	u	m	1	m	u	1	2	2
be	d	e	p	e	**ri**	u	u	u	0	m	m	2	2	4
	dr	e	**mp**	_		m	m	m	3	u	u	0	3	1
be	**dr**	e	**mp**	e		m	m	u	2	m	u	1	3	3
	dr	e	**mp**	e	**ri**	m	m	u	2	u	m	1	3	3
be	**dr**	e	p	_		m	u	m	2	m	u	1	3	2
be	d	e	**mp**	_		u	m	m	2	m	u	1	3	2
be	**dr**	e	p	e	**ri**	m	u	u	1	m	m	2	3	4
be	d	e	**mp**	e	**ri**	u	m	u	1	m	m	2	3	4
be	**dr**	e	**mp**	_		m	m	m	3	m	u	1	4	2
be	**dr**	e	**mp**	e	**ri**	m	m	u	2	m	m	2	4	4

Table 3: *Number of non-words in TOPhS with each combination of number of syllables and marked structures*

		Number of syllables				
		1	**2**	**3**	**4**	**Totals**
Number of marked structures	**0**	0	4	0	0	4
	1	4	8	8	0	20
	2	8	8	16	4	36
	3	4	8	8	8	28
	4	0	4	0	4	8
	Totals	16	32	32	16	96

Table 4: *Mean scores (SD) of the two SLI groups and their comparison*

		SLI-high	**SLI-low**	**p-value**	**d**
	TOPhS TNC (/94)	79.4 (9.3)	38.7 (6.2)	**<.001**	**5.08**
	TOPhS PPC	.97 (0.02)	.85 (0.05)	**<.001**	**3.24**
	VATT total (/40)	32.1 (6.2)	9.6 (6.8)	**<.001**	**3.48**
	VATT agreement (/20)	16.8 (2.7)	3.6 (4.9)	**<.001**	**3.40**
	VATT tense (/20)	15.1 (4.0)	6.0 (3.4)	**<.001**	**2.63**
	Age (in months)	155.3 (15.2)	159.1 (15.6)	.63	-0.25
	BPVS raw score	95.0 (16.0)	86.6 (11.4)	.27	0.60
[a]	TROG raw score	15.6 (2.8)	15.1 (1.9)	.44	
[a]	CELF Receptive z-score	-2.06 (0.35)	-2.21 (0.29)	.42	
	Concepts & Directions raw score	17.7 (7.2)	13.9 (5.8)	.28	0.58
	Word Classes raw score	19.9 (2.2)	17.1 (6.5)	.33	0.60
[a]	Semantic Relationships raw score	8.8 (4.8)	10.6 (6.1)	.96	
[a]	CELF Expressive z-score	-2.20 (0.34)	-2.19 (0.35)	.88	
[a]	Formulated Sentences raw score	22.4 (7.0)	17.7 (7.4)	.10	
[a]	Recalling Sentences raw score	26.0 (10.7)	17.7 (14.8)	.09	
	Sentence Assembly raw score	9.9 (3.8)	10.4 (4.1)	.79	0.13
	BAS-II z-score	0.03 (1.96)	-0.13 (0.68)	.70	0.19
	WORD Reading Accuracy raw score	30.9 (12.6)	24.4 (4.9)	.23	0.71
	WORD Spelling Accuracy raw score	27.1 (8.0)	22.7 (5.3)	.24	0.69
	WORD Reading Accuracy z-score	76.5 (17.6)	66.6 (5.5)	.17	0.74
	WORD Spelling Accuracy z-score	79.8 (14.4)	70.1 (8.4)	.15	0.80

[a] non-parametric test carried out due to non-normal distributions

Table 5: *Mean scores (SD) and ranges on the VATT: agreement, tense and comparisons of performance within and across groups.*

	Agreement		**Tense**		**Agr vs. Tns**
Group	**Mean (SD)**	**Range**	**Mean (SD)**	**Range**	**p-value**
SLI-low	3.6 (4.9)	0-14	6.0 (3.4)	0-10	.03
LC-low	19.2 (0.8)	19-20	17.9 (2.6)	12-20	<.001
p-values					
SLI-low vs. LC-low	<.001		<.001		
SLI-high	16.6 (2.8)	12-20	14.4 (3.7)	7-18	.09
LC-high	17.1 (5.0)	4-20	17.0 (3.7)	10-20	.84
AC-high	20.0 (0.0)	20-20	19.4 (0.5)	19-20	.11
p-values:					
all groups	.03		.006		
SLI-high vs LC-high	.35		.03		
SLI-high vs AC-high	.01		<.001		

Values differ slightly from Table 4 as data were missing for both language and age controls for on SLI-high child, whose data were excluded from this table and statistical analyses

Table 6: *Correlation coefficients of TOPhS TNC score with other measures. Partial correlations for controls partial out age.*

	SLI	**Controls**	
	Spearman r	Pearson r	Partial correlations
Age	-.003	**.31***	-
TOPhS PPC	**.94*****	**.93*****	**.93*****
VATT (total)	**.91*****	.28	.27
VATT (agreement)	**.97*****	**.43***	.16
VATT (tense)	**.89*****	**.39***	.30
BPVS raw score	.44	**.42****	**.32***
TROG raw score	.41	**.30***	.13
Reading Accuracy raw score	.26	n/a	n/a
Spelling Accuracy raw score	.41	n/a	n/a

p-values (1-tailed): p<0.05*, p<0.01**, p<0.001***

Table 7: *Spearman rank correlations (r) and partial correlations for Length, Metrical complexity and Added consonants, partialling out the other two factors.*

	Spearman r			**r when other two partialled out**		
	SLI-low	**SLI-high**	**Controls**	**SLI-low**	**SLI-high**	**Controls**
Correct / incorrect scoring						
Length	**-.64*****	**-.46*****	**-.33*****	-.10	+.03	+.14
Phonological complexity	**-.61*****	**-.51*****	**-.41*****			
Metrical complexity	**-.71*****	**-.56*****	**-.46*****	**-.38*****	**-.32****	**-.35*****
Syllablic complexity	-.14	-.16	-.10			
Added consonants	**-.39*****	**-.32*****	-.17	**-.49*****	**-.36*****	-.14
Word End	**+.28****	+.13	<-.01			
Percentage Phonemes Correct scoring						
Length	**-.38*****	**-.23***	-.02	+.15	+.02	+.08
Phonological complexity	**-.60*****	**-.39*****	-.06			
Metrical complexity	**-.49*****	**-.27****	-.04	**-.38*****	**-.19***	-.10
Syllablic complexity	**-.30****	**-.24***	-.05			
Added consonants	**-.45*****	**-.37*****	-.07	**-.50*****	**-.33****	-.08
Word End	+.06	+.04	-.05			

p-values (1-tailed): p<0.05*, p<0.01**, p<0.001***

Table 8: *Summary of findings*

Test	SLI low vs high	Comparison with controls: SLI-low	Comparison with controls: SLI-high	Correlated with TOPhS?: SLI	Correlated with TOPhS?: Controls
TOPhS	low < high	< Lang, < Age	= Lang, = Age	N/A	N/A
TOPhS PPC	low < high	< Lang, < Age	= Lang, = Age	yes ++	yes ++
VATT tense	low < high	< Lang, < Age	< Lang, < Age	yes ++	no
VATT agreement	low < high	< Lang, < Age	= Lang, < Age	yes ++	no
BPVS	low = high	= Lang, < Age	= Lang, < Age	no	yes
TROG	low = high	= Lang, < Age	= Lang, < Age	no	no
WORD Reading Accuracy	low = high	<Age (from z-score)	<Age (from z-score)	no	
WORD Spelling Accuracy	low = high	<Age (from z-score)	<Age (from z-score)	no	
CELF-3 Receptive Language	low = high	<Age (from z-score)	<Age (from z-score)		
CELF-3 Expressive Language	low = high	<Age (from z-score)	<Age (from z-score)		
BAS-II z-score	low = high	=Age (from z-score)	=Age (from z-score)		

Lang = language controls, Age = age controls

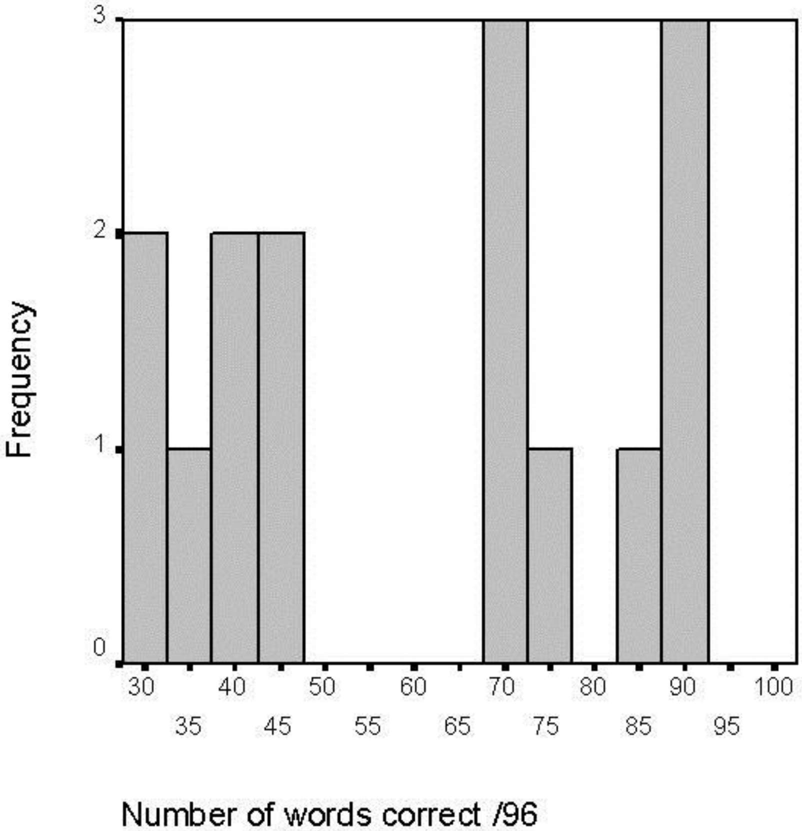

Figure 1: Histogram showing bi-modal distribution of SLI scores on TOPhS test

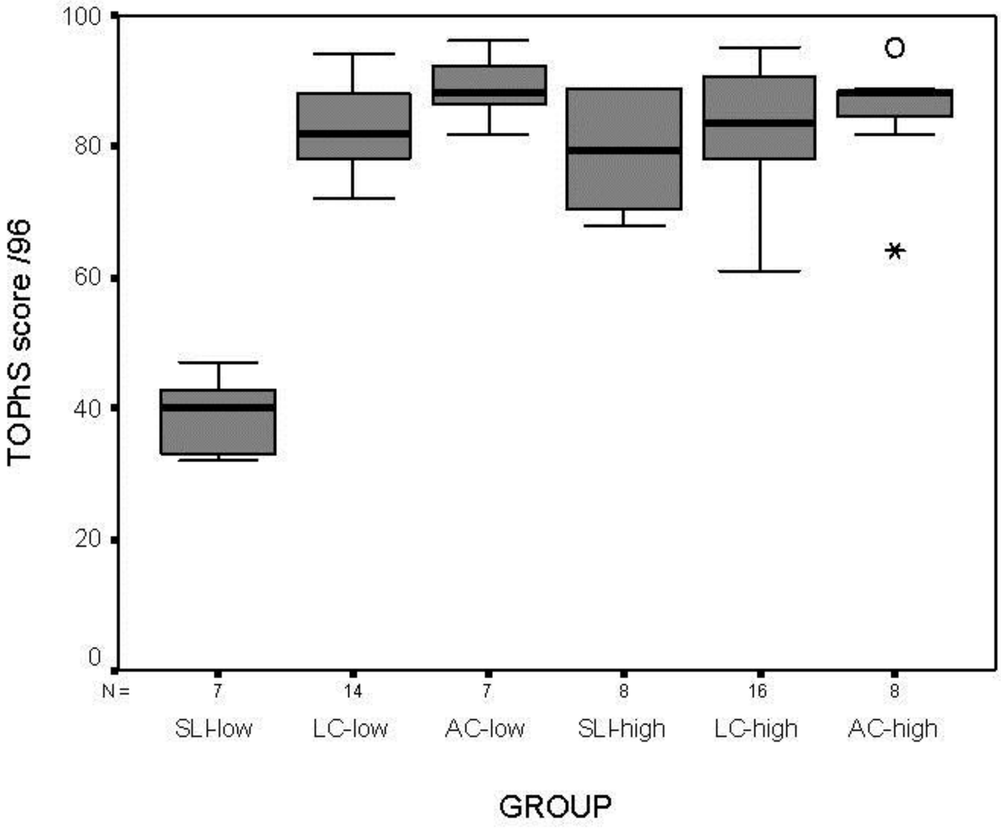

Figure 2: Boxplot showing distribution of TOPhS scores (/96) for SLI-low and SLI-high groups, their language and age controls

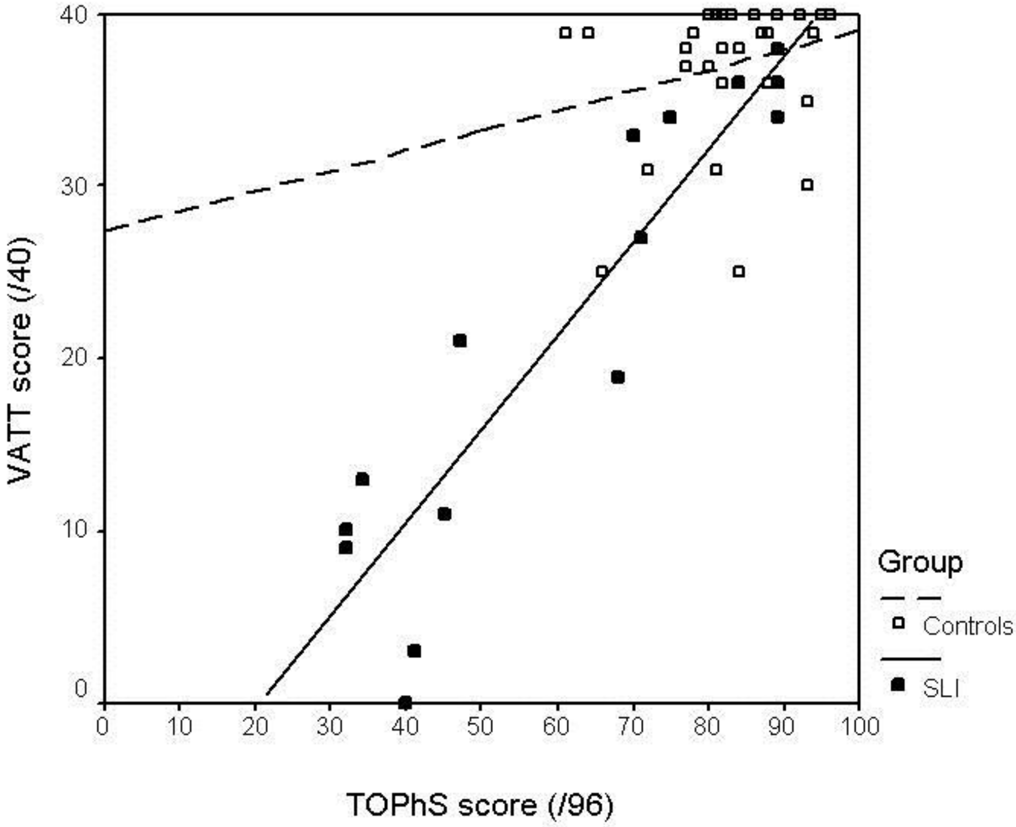

Figure 3: The relationship between TOPhS and VATT for the participants with SLI versus controls

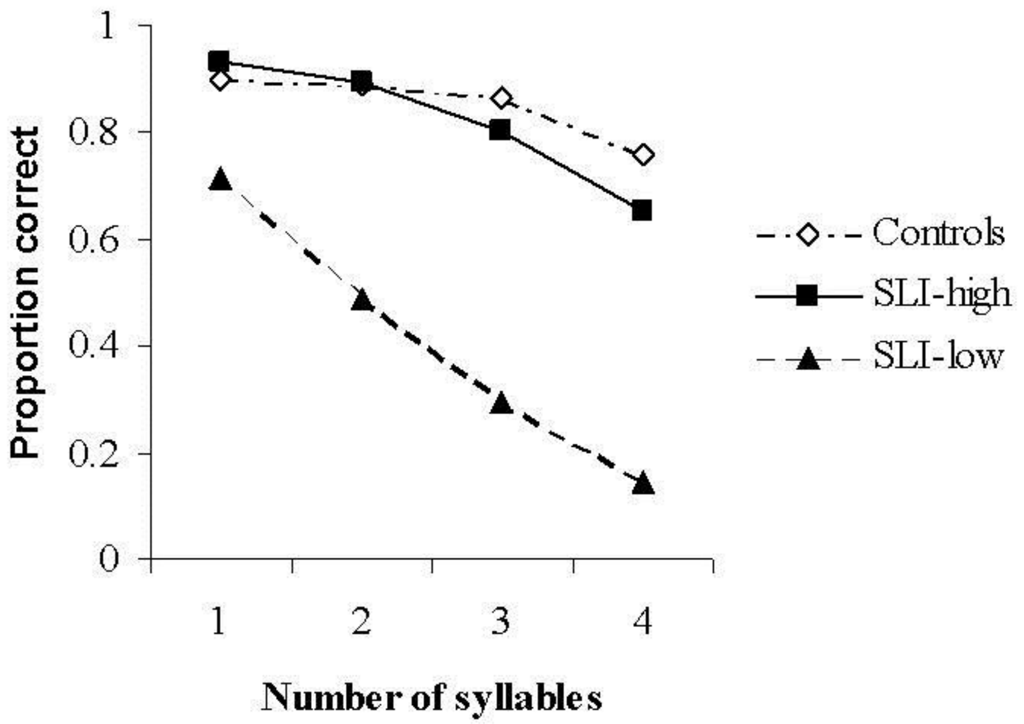

Figure 4: The influence of length (number of syllables)

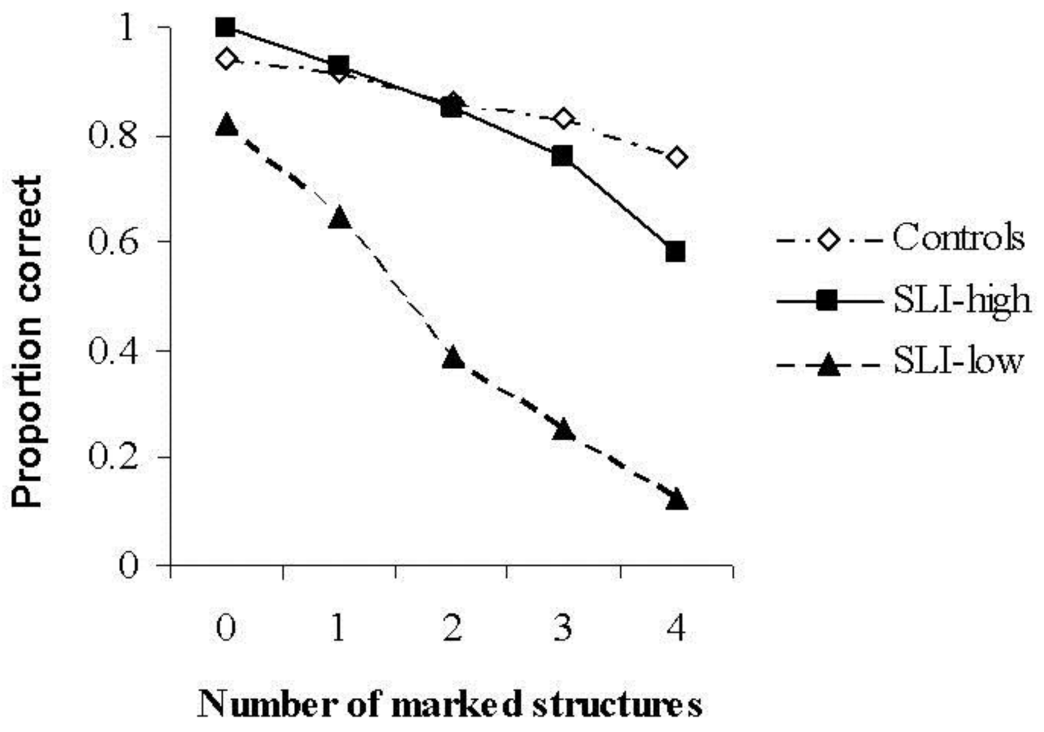

Figure 5: The influence of phonological complexity (number of marked structures)