Non-word repetition in adolescents with Specific Language Impairment and Autism plus Language Impairments: a qualitative analysis. Riches NG, Loucas T, Baird G, Charman T, Simonoff E. J Commun Disord. 2011 Jan-Feb;44(1):23-36. Epub 2010 Jul 8. PMID: 2067391

Non-word repetition in adolescents with Specific Language Impairment and Autism plus Language Impairments; a qualitative analysis

Riches, N.G., Loucas, T., Baird, G., Charman, T. & Simonoff, E.

Abstract

Non-word repetition (NWR) was investigated in adolescents with typical development, Specific Language Impairment (SLI) and Autism plus Language Impairment (ALI) (n = 17, 13, 16, and mean age 14;4, 15;4, 14;8 respectively). The study evaluated hypothesis that poor NWR performance in both groups indicates an overlapping language phenotype (Kjelgaard & Tager-Flusberg, 2001). Performance was investigated both quantitatively, e.g. overall error rates, and qualitatively, e.g. effect of length on repetition, proportion of errors affecting phonological structure, and proportion of consonant substitutions involving manner changes. Findings were consistent with previous research (Whitehouse et al. 2008) demonstrating a greater effect of length in the SLI group than the ALI group, which may be due to greater short-term memory limitations. In addition, an automated count of phoneme errors identified poorer performance in the SLI group than the TD group. These findings undermine the overlapping phenotype hypothesis. Errors affecting phonological structure were relatively frequent, accounting for around 40% of phonemic errors, but less frequent than straight consonant-for-consonant or vowel-for-vowel substitutions. It is proposed that these two different types of errors may reflect separate contributory mechanisms. Around 50% of consonant substitutions in the clinical groups involved manner changes, suggesting poor auditory-perceptual encoding. From a clinical perspective algorithms which automatically count phoneme errors may enhance sensitivity of NWR as a diagnostic marker of language impairment.

Literature Review

Over the past few years there has been vigorous debate over the relationship between Specific Language Impairment (SLI) and Autism Spectrum Disorders (ASD) (Williams, Botting, & Boucher, 2008). While standard diagnostic criteria differentiate these two disorders, recent clinical data suggest that the boundaries may not be so clear. According to the Diagnostic and Statistical Manual of the American Psychiatric Association (DSM IV: 2000) children with SLI should not present with a Pervasive Developmental Disorder of any kind. While the presence of SLI does not act as an exclusionary criterion for ASD, spoken language delay is but one of four communicative traits, including stereotyped speech, poor conversational ability and lack of imaginative play, and an individual may be diagnosed with ASD if only one of these traits is present. Moreover, there is no mention in DSM IV of structural, i.e. syntactic, language difficulties in ASD. Nonetheless, despite the tendency to view SLI and ASD as separate disorders, recent studies have identified a group of individuals with ASD *and* poor language abilities, henceforth referred to as "ALI"; Autism plus language Impairment. These children perform poorly on clinical markers of SLI such as past tense elicitation tasks and nonword repetition (Kjelgaard & Tager-Flusberg, 2001; Roberts, Rice, & Tager-Flusberg, 2004). Moreover, like children with SLI they also exhibit a language profile whereby syntax is more severely affected than lexical abilities and present with a dissocation between verbal and non-verbal abilities (Kjelgaard & Tager-Flusberg, 2001; Tager-Flusberg & Joseph, 2003). Tager-Flusberg and colleagues argue that similar performance across a range of behavioural measures, e.g. nonword repetition and past tense tasks, suggests a phenotypic overlap between these two groups. A phenotype is a manifestation of an underlying genetic code, and therefore the implication is that individuals with SLI and ALI may share *at least some* genetic material. A possible genetic link between SLI and ALI has been investigated in a number of studies using behavioural measures and laboratory-based genetic sequencing (see Williams et al., 2008 for a critical evaluation).

Nonword repetition (NWR), has played a key role the debate on phenotypic overlap. This simple task involves encouraging children to repeat nonsense words. Errors are counted and a total score is calculated. NWR is a promising phenotypic marker of SLI because it shows high rates of sensitivity and specificity (Conti-Ramsden, Botting, & Faragher, 2001), and it is also highly heritable (Bishop, North, & Donlan, 1996), in keeping with the claim that SLI has a strong genetic basis. In young children, it also correlates closely with assessments of vocabulary (Gathercole, Willis, Emslie, & Baddeley, 1992), Mean Length of Utterance (MLU), and morphosyntactic complexity (Adams & Gathercole, 1996, 1995), suggesting that it engages cognitive mechanisms which underlie general language abilities. Nonetheless, it is also apparent that NWR is a cognitively complex task, and it is difficult to determine where it may break down. While many have argued that NWR depends on the phonological loop, a short-term memory system, it is also clear that NWR is influenced by factors such as wordlikeness and phonotactics (Munson, Kurtz, & Winsor, 2005), and therefore, to an extent must be influenced by phonological representations in long-term memory (LTM). Furthermore, repetition is obviously dependent on the children hearing and encoding the stimulus correctly, and therefore may be influenced by auditory processing and phonetic encoding difficulties. Given that NWR is a complex task, which may break down at a variety of levels, it clear that different populations may have similar NWR performance due to different underlying cognitive factors.

A recent study of the NWR in children with ALI and SLI, mean ages 11;10 and 10;11 respectively, (Whitehouse, Barry, & Bishop, 2008) has closely scrutinized the claims of the overlapping phenotype hypothesis by qualitatively investigating NWR errors in order to tease apart underlying mechanisms. There was a qualitative difference between the SLI and ALI groups such that significant differences were observed for the five-syllable nonwords, with a trend towards a significant effect for four-syllable nonwords. In this way the SLI group displayed a more concave profile characterised by a sudden increase in error rates for words of four and five syllables. A two-way ANOVA found a significant interaction between group, and length, thereby statistically confirming that profiles differed across groups. Whitehouse et al. (2008), argue that while the large effect of length in the SLI group is usually ascribed to short-term memory (STM) limitations, the pattern in the ALI group suggests that STM plays a lesser role, and therefore poor performance may be due to a different causal factor, for example difficulties orienting attention towards speech stimuli (e.g. Ceponiene et al., 2003). In this way a qualitatively distinct error profile implies a different causal mechanism, which in turn undermines the phenotypic overlap hypothesis.

The analysis of Whitehouse et al. (2008) are clearly in need of replication and this is one aim of the current study. In addition, further qualitative analyses of NWR performance will be conducted in order to investigate qualitative differences. One analysis will assess the degree to which errors are structure-changing or structure-preserving. For example, the following error; /<u>bl</u>ɒn.tə.st<u>eɪ</u>. pɪŋ/ → /<u>l</u>ɒn.tə.st<u>e</u>.pɪŋ/ is structure-changing in that it involves the simplification of the initial onset, and the penultimate nucleus. According to Gallon, Harris and van der Lely (2006), structure changing errors such as reduction of onset clusters, and weak syllable deletion may be characteristic of children with Grammatical Specific Language Impairment (G-SLI), i.e. children who perform poorly on a wide range of assessments which focus on the use of grammatical structures. However, by contrast, Marton and Schwartz (K. Marton & Schwartz, 2003; Klara Marton, 2006) have argued that NWR errors made by children with "general" SLI, who have been diagnosed with omnibus assessments, tend to be structure-preserving, with majority of errors being Consonant-for-Consonant (C-for-C) substitution errors, particularly substitution of consonants. The issue of whether an error is structure-changing or structure-preserving is critical as it can elucidate the origin of NWR difficulties. Structure-changing errors may reflect difficulties with hierarchical prosodic and phonological processes (see Gallon et al., 2006, for a discussion). By contrast, structure-preserving errors may arise from difficulties with simultaneous processing of metrical information (number of syllables, and stress placement), and melodic information (phonemic features), which results in preservation of the former, and disruption to the latter (K. Marton & Schwartz, 2003).

An additional qualitative analysis will focus on consonant substitutions. Shriberg et al. (2009) argue that the nature of these substitutions can elucidate difficulties with auditory perceptual encoding. They distinguish between "within-class" substitutions, where "class" refers to *manner* of articulation and "between-category" substitutions, involving manner changes. /t/ → /d/ is a within-class substitution, preserving plosive manner but changing voice, while /t/ → /tʃ/ is a between-category substitution, changing manner from plosive to affricate. The authors argue that a substitution preserving manner can be interpreted as a partial encoding of the target, i.e. a near miss. In studies of phonemic perception, manner changes are more likely to result in the perception of distinct phonemes, than changes in place or voicing (e.g. Bailey & Hahn, 2005), suggesting that spectral characteristics linked to manner provide a strong cue for phoneme identification. Therefore, individuals who make frequent manner changes during repetition have overridden a strong perceptual cue, and this behavior may reflect poor auditory-perceptual encoding. Shriberg et al. (2009) found that in typically developing children, aged 3;0 to 5;0 the percentage of consonant substitutions involving manner changes was significantly lower than age matched groups of children with speech delay, language-impairment, and language-impairment plus speech delay, suggesting auditory-perceptual encoding difficulties in the latter groups.

In addition to conducting a range of qualitative analyses this study is novel in that it investigates NWR in adolescents. Follow-up studies of children with SLI into adolescence have found that while they may improve on a variety of standardized assessments of spoken and written language, their performance on NWR remains poor compared to age-matched peers (Stothard, Snowling, Bishop, Chipchase, & Kaplan, 1998). Therefore, it may be argued that NWR difficulties after childhood reflect a persistent cognitive difficulty which is likely to be a core component of NWR difficulties in individuals irrespective of their age. By contrast, NWR in childhood may be confounded by difficulties in other areas. For example, Speech Sound Difficulties, which may affect NWR, are relatively common, but often resolve in a short period (Vance, Stackhouse, & Wells, 2005). The instability of NWR in early childhood has been demonstrated by one study which tracked children who had performed poorly on NWR and other STM tasks aged 4;0 over a 4-year period (Gathercole, Tiffany, Briscoe, Thorn, & ALSPAC, 2005). Just under half of these children (38%) recovered from their early STM difficulties. Given the instability of NWR in childhood, data from adolescents may be important in order to identify the core mechanisms involved in NWR performance.

Another novel characteristic of the present study is that it employs an automatic algorithm to code errors, the Levenshtein Distance (Levenshtein, 1966). This counts the minimum number of operations required to transform string A into string B, where operations are addition, substitution or omission. For example, the following transformation; *chalk* → *cheese* yields an LD of 4; 3 substitutions (*a* → *e*, *e* → *l*, *l* → *s*) and 1 addition (*e*). While the Children's Test of Nonword Repetition (Gathercole & Baddeley, 1996), one of the most widely used NWR assessments, counts errors in an all-or-none fashion, scoring the word as correct if it contains no errors, and incorrect if it contains one or more error, the LD clearly has an advantage in that it distinguishes between words with one error and words with one or more error. For example, a participant who repeats words relatively accurately, but makes frequent voicing errors, e.g. *dopelate* → *topelate*, might, under the all-or-none scheme, obtain the same score as a participant with more severe repetition difficulties affecting multiple phonemes per word. The LD is capable of distinguishing between these two participants, and we suggest that it is therefore a more sensitive metric of repetition difficulty.

In summary, the following null hypotheses, are proposed;

- Overall NWR error rates will be statistically equivalent in the ALI and SLI groups.
- Qualitatively there will be no difference between the groups in terms of (a) the effect of length on error rates, (b) the proportion of errors affecting syllable structure, and (c) the percentage of consonant substitutions involving manner changes

Any quantitative or qualitative differences between groups will be interpreted as undermining the phenotypic overlap hypothesis.

Methodology

Participants

16 participants with ALI, and 11 participants with SLI were selected from a cohort of individuals with Special Educational Needs who had been assessed during the Special Needs and Autism (SNAP) Project (Baird et al., 2006). A diagnosis of autism was made on the basis of ICD-10 criteria (World Health Organisation, 1993) using information from the Autism Diagnostic Observation Schedule (ADOS: Lord et al., 2000), Autism Diagnostic Inventory - Revised (ADI-R: Lord, Rutter, & Couteur, 1994), clinical vignettes and teacher report. Participants were diagnosed with language impairment if there was a discrepancy between their language abilities, measured using the Clinical Evaluation of Language

Fundamentals -3 UK (CELF-3 UK: Semel, Wiig, & Secord, 2000), and their non-verbal IQ scores on the Wechsler Intelligence Scale for Children-3UK (Wechsler, 1992). The language cut-off was a standard score of 77 or below on the expressive and/or receptive subscales (z = -1.55), while the IQ cut-off was a standard score of 80 or above on either Performance IQ, or the Perceptual Organisational Index (z = -1.35). No individual met diagnostic criteria for any syndrome other than ASD or SLI according to teacher report. All participants were tested for hearing difficulties (<30dB).

Given the time lag between the SNAP study and the current study, on average 42 months, language and non-verbal abilities were retested using a shorter version of the previous assessments; Concepts and Directions (CD) and Recalling Sentences (RS) from the CELF, and Picture Arrangement (PA) and Block Design (BD) from the WISC. The two CELF subtests were chosen to measure expressive (RS) and receptive (CD) abilities, with the former being an especially reliable indicator of SLI (Conti-Ramsden et al., 2001).

Two further participants with SLI were recruited via contacts in schools with language units. Non-ASD status was determined using the ADOS (participant 1), SCQ (both participants) and the ADI-R (participant 2). Language and non-verbal abilities were assessed using the WISC and the CELF, with the full battery used for participant 1, and the reduced battery used for participant 2. Hearing difficulties were assessed via teacher report.

A further 17 typically-developing children (TD) were recruited from a single school in South London. These children were screened for ASD using the SCQ (all scores below 7). Verbal and non-verbal abilities were determined using the short versions of the CELF and WISC.

Descriptives are shown in Table 1. WISC scores did not vary significantly across the groups indicating similar nonverbal abilities ($p = .854$, full analyses shown in table). Age did vary significantly ($p < 0.001$**) with Tukey's tests finding the following significant contrasts; TD < SLI, TD < ALI. The participants with SLI were therefore significantly older than the other groups. A number of assessments of STM / Working Memory (WM) from the Working Memory Test Battery for Children (WMTB-C: Gathercole & Pickering, 2001) were administered including Digit Recall (DR), Backwards Digit Recall (BDR), and Listening Recall (LR), a version of the listening span task where children must listen to a block of sentences, make a true / false judgment after each sentence, and at the end of the block recall all of the final words in the right order. The participants with ALI were better on average at the short-term memory tests than the participants with SLI, obtaining significantly higher scores on the Digit Recall task ($p = 0.025$*).

Stimuli

Children were administered two separate assessments of non-word repetition; the Children's Test of Non-word Repetition (CNRep: Gathercole & Baddeley, 1996) and the NonWord Memory Test (NMT: Gathercole & Baddeley, 1989). The purpose of combining the two assessments was to increase the number of stimuli, and henceforth the power of the study. Each assessment comprises a pre-recorded set of nonwords, 40 in the former, and 28 in the latter. The stimuli were designed to be phonotoctically similar to real English words, and therefore contain complex onsets and nuclei, e.g. consonant clusters (*glistering),* long vowels and diphthongs (*comeecitate).* Some of the words contained derivational morphemes, e.g. –*ing* and –*ate* above. While such derivational morphemes entail that representations in LTM can be recruited to support maintenance and rehearsal, they nonetheless enhance the wordlikeness of the stimuli. While some of the stimuli contained real English words, these tended to be low frequency, e.g. *brew* in *brufid*. 5 stimuli; *peneriful, empliforvent, perplisteronk, frescovent,* and *brasterer*, were repeated across the two assessments so children's responses for these words in the NMT were excluded from the analysis. This process affected the counterbalancing of words by syllable length so that there were slightly more words in the low syllable groups (17 words of 2 syllables, 17 words of 3 syllables, 15 words of 4 syllables, 14 words of 5 syllables). However, all analyses investigate the mean number of errors per word, so differences in the size of the syllable groups have a minimal impact on the findings.

Procedure

The test was administered using the original spoken recordings and were played to the children over headphones. Original test procedures, e.g. the use of practice items, were used to familiarize the children with task. Children's responses were recorded using the inbuilt microphone on the laptop (Sigma-Tel C Major Audio) and were stored as .wav files.

Coding

Transcription methods

Children's responses were coded directly from the .wav files by the first author using broad phonetic transcription. International Phonemic Association (IPA) symbols were transformed into standard ASCII characters for the purpose of data analysis. Characters were chosen which closely resembled their IPA counterparts, e.g. /t/ was transcribed as *t*. However, it was not always possible to find a closely matching ASCII character, e.g. /θ/ was transcribed as *T*. Affricates, e.g. /dʒ/ in *judge*, were regarded as monophones, given that they are in complementary distribution with other monophones, e.g. /dʒæm/ versus /dæm/. They were therefore transcribed using a single character. Long vowels, e.g. /i:/ were also coded with a single character despite occupying two time slots. This is because, in English, changes in vowel length also lead to changes in place of articulation, e.g. /ʃi:p/ → /ʃɪp/. If long vowels had been represented by two characters this would have inflated the error count involved in vowel reduction, e.g. /ʃi:p/ → /ʃɪp/ involves one character substitution (i → ɪ) and one character addition (:). This coding scheme would have excessively penalized an error which depends heavily on an individuals' perception of vowel length and is therefore difficult to code reliably. Diphthongs, by contrast were coded using two characters.

In addition to the phonemic coding scheme described above, a CV coding scheme was used to capture changes in syllable structure. This scheme merely replaced all consonants with C and all syllables with V. According to the CV coding scheme, long vowels were represented by two characters; VV, given that the coding scheme is designed to investigate changes to phonological structure, and therefore length is a primary variable of interest.

Calculation of error rates

The Levenshtein Distance between the stimulus and responses was calculated for both the phonemic transcriptions and the CV transcriptions. For the phonemic coding scheme, two types of error rates were calculated. To calculate to the "all-or-none" error rate, repetitions were scored as 1 if the response yielded an LD of one or more, or 0 if the participant made no errors. This coding scheme is identical to that used by the CNRep. The "phonemic" error rate was merely the LD, i.e. the number of single-phoneme substitutions, additions or omissions required to transform the stimulus into the response.

For the CV scheme, error rates were also calculated using the LD. This effectively counted the number of single-phoneme substitutions, additions or omissions which altered syllabic structure by changing the basic CV structure of the word. A number of syllabic errors could be identified using this scoring method; the addition or omission of consonants within onsets, the addition of a consonant within the nucleus to create a coda, vowel reduction or vowel-lengthening within the nucleus, and addition / omission of any phonemes resulting in changes to the number of syllables in the word. The only type of error excluded from this error count was within-category substitutions, e.g. C for C or V for V.

By comparing the LD for the phonemic and syllabic coding schemes we can effectively determine the proportion of operations which affected syllabic structure. For example, for the following transformation; *dopelate* /dɒ.pə.leɪt/ → *doslate* /dɒs.leɪt/, the phonemic coding scheme yields two errors; *p* is substituted by *s*, and *e* is omitted. However, only one of these errors, the omission of *e*, affects syllable structure. This is reflected in the CV coding scheme; CV.CV.CVVC → CVC.CVVC, resulting in one error; the omission of the second V. Therefore by dividing phonemic error rates by CV error rates, we can calculate the proportion of error operations which affect syllable structure, in this case 50%.

Identifying consonant substitutions

The final analysis investigated consonant substitutions, which were coded using the ASCII transcripts. A consonant was deemed to be substituted if it was flanked by phonemes which did not vary between the stimulus and the response, e.g. dɒ**p**əleɪt → dɒ**t**əleɪt. Word boundaries were ignored, so that changes in initial or final consonants were deemed to be substitutions according to whether they were next to a single unchanging phoneme, e.g. dɒpəleɪt → tɒpəleɪn.

Reliability

The second author transcribed the .wav files for 2 TD participants, 2 with SLI and 2 with ALI, comprising 13% of the dataset. His transcriptions were compared against those of the first author. Disagreements arose on 4% of items, yielding an agreement rate of 96%

Analysis

Analysis of error rates based on phonemic coding scheme

Error rates by Group are reported in Table 2, and plotted in Figure 1. A Levene's test comparing the SLI and ALI groups on all-or-none error rates yielded a non-significant pvalue ($p = .338$) thereby indicating homogeneity of variance. A one-way ANOVA was run investigating the effect of Group on error rates derived from the phonemic transcription, followed by a post hoc Tukey's test. There was a significant effect of Group on all-or-none error rate ($F(2, 43) = 13.977$, $p < 0.001$**, partial $\eta^2 = 0.394$), with Tukey's test revealing the following significant contrasts; TD < SLI, TD < ALI, but no difference between the clinical groups (SLI=ALI). Likewise there was a significant effect of Group on phonemic error rate ($F(2, 43) = 12.327$, $p < 0.001$*, partial $\eta^2 = 0.364$), with Tukey's test revealing the following significant contrasts; TD < SLI, TD < ALI, ALI < SLI.

Further analyses investigated differences between the clinical groups. ANOVAs were conducted to investigate the interaction between Group (SLI versus ALI), and syllable length (2, 3, 4, 5). For all-or-none error rates there was a trend towards a significant effect of Group ($F(1, 27) = 3.013$, $p = 0.094$, partial $\eta^2 = 0.121$), a significant effect of Length ($F(3, 81) = 34.534$, $p < 0.001$**, partial $\eta^2 = 0.561$), but no significant interaction ($F(3, 81) = 1.470$, $p = 0.229$, partial $\eta^2 = 0.052$). For phonemic error rates there was a marginally significant effect of Group ($F(1, 27) = 4.226$, $p = 0.050$, partial $\eta^2 = 0.151$), a significant effect of Length ($F(3, 81) = 28.556$, $p < 0.001$**, partial $\eta^2 = 0.514$), and moreover, a significant interaction ($F(3, 81) = 3.262$, $p = 0.026$*, partial $\eta^2 = 0.108$). Oneway ANOVAs investigated the effect of Group on error rates for each word length. There was a significant difference between groups for the four-syllable words ($F(1, 27) = 5.823$, $p = 0.023$*, partial $\eta^2 = 0.177$), a trend towards a significant effect for five-syllable words ($F(1, 27) = 3.491$, $p = 0.073$, partial $\eta^2 = 0.115$), and non-significant differences for words of two and three syllables ($F(1, 27) = 0.206$, $p = 0.654$, partial $\eta^2 = 0.008$ and $F(1, 27) = 1.274$, $p = 0.269$, partial $\eta^2 = 0.045$ respectively).

Analysis of error rates based on CV coding scheme

The mean LD for the CV coding scheme is shown by the lighter bars in Figure 2. There was a significant effect of Group on these syllabic errors $F(2, 43) = 8.992$, $p = 0.001$, partial $\eta^2 = 0.295$), with post-hoc Tukey's test revealing only a significant difference between the TD and SLI groups (TD < SLI, SLI = ALI). By comparing the LD for the syllabic transcriptions with the LD for the phonemic transcription, we can ascertain what proportion of phoneme operations also affected syllabic structure. For each group the LD for the CV coding scheme was divided by the LD for the phonemic coding scheme and multiplied by 100, which gives the percentage of phoneme errors affecting syllable structure. The SLI participants produced, on average, a larger proportion of errors affecting syllabic structure than the ALI group (SLI: mean = 44.3%, s.d. = 12.8, ALI: mean = 37.9%, s.d. = 12.7). However, there was no effect of Group on the proportions calculated on a participant-by-participant basis ($t(27) = 1.346$, $p = 0.190$, $d = 0.042$). Changes affecting the number of syllables were also calculated. Responses which reduced the number of syllables were rare in both groups (2.2% in the SLI group, and 0.4% in the ALI group), and there were no significant effect of Group on the number of syllable reductions calculated on a participant-by-participant basis ($t(27) = 1.346$, $p = 0.190$, $d = 0.042$). Errors involving added syllables were extremely rare in both groups, at less than 1%.

Finally the interaction between Group and Length was investigated with proportion of errors affecting syllable structure as the dependent variable. There was no effect of Group ($F(1, 27) = 2.807$, $p = 0.105$, partial $\eta^2 = 0.042$), a significant effect of Number ($F(3, 80) = 3.308$, $p = 0.024$*, partial $\eta^2 = 0.110$), and no significant interaction ($F(3, 80) = 1.257$, $p = 0.295$, partial $\eta^2 = 0.045$). The significant effect of Number was driven by the trend for a lower proportion of structure-changing errors in the longer words, which changed from 50% to 41% to 34% to 35% with each extra syllable, with data collapsed across groups.

Analyses of consonant substitutions

The language impaired groups made more repetition attempts involving at least one consonant substitution. Overall 9.9% of the responses of the TD children involved 1 or more consonant substitution (s.d. 29), compared to 20.3% in the SLI group (s.d. = 40) and 17.7% in the ALI group (s.d. = 38). Large standard deviations reflect wide variation within groups. There was a significant effect of Group on the percentage of repetition attempts involving at least one consonant substitution ($F(2, 43) = 13.619$, $p < 0.001$**, partial $\eta^2 = 0.388$), with follow-up Tukey’s tests indicating the following contrasts; TD < SLI, TD < ALI, SLI = ALI. The proportion of consonant substitutions involving changes of manner was calculated. This proportion was marginally lower in the TD group with 40.7% of all consonant substitutions involving changes of manner (s.d. = 25.4), compared to 51.0% in the SLI group (s.d. = 9.9), and 54% in the ALI group (s.d. = 15.3). A one-way ANOVA with Group as the independent variable, and proportion of substitutions involving manner changes for each participant as the dependent variable, found no effect of Group ($F(2, 43) = 1.518$, $p = 0.231$, partial $\eta^2 = 0.066$)

Discussion

The study replicated and extended the findings of Whitehouse et al. (2008), with a significantly stronger effect of syllable length in the SLI group than the ALI group. In addition, when errors were counted on a phoneme-by-phoneme basis, the participants with SLI performed significantly poorer overall than the ALI participants. Further qualitative analyses did not uncover other robust differences between the clinical groups. While the participants with SLI tended to make more errors affecting syllable structure, and tended to drop slightly more syllables, there were no significant differences between the groups in the proportions of such errors. Overall, children in both clinical groups tended to preserve the number of syllables with just over half of their errors consisting of C-for-C or V-for-V substitutions, consistent with the structure-preservation account (K. Marton & Schwartz, 2003). However, errors affecting syllabic structure, i.e. the addition or omission of material within onsets or nuclei, were also common, in line with the structure-changing pattern witnessed by Gallon et al. (2006). Consonant substitutions were also analysed in order to determine the role of perceptual difficulties. Manner changes were proportionally more frequent in the clinical groups providing some support for a phonemic encoding account as proposed by Shriberg et al. (2009), though between-group differences were not significant.

The most significant finding of the study is the interaction between Group and Length, also observed by Whitehouse et al. (2008). This pattern suggests a more severe limitation of Phonological STM in the SLI group. Significant between-group differences were observed on the three-syllable nonwords, with a trend towards significance on four-syllable non-words, a pattern similar to that observed by Whitehouse et al. (2008) who also found significant or near-significant differences on the longer words. This pattern suggests a more concave profile in the SLI group with a sudden decrease in performance for words greater than three syllables, a pattern which suggests that individuals with SLI have a limited capacity Phonological STM such that when a nonword exceeds a certain length recall is severely affected.

In addition to the NWR data, a wide range of additional assessments involving STM and WM were conducted. The participants with SLI performed worse on all them. Performance was particularly poor on Digit Recall, a task designed to assess STM. More complex assessments with a stronger WM component, such as Listening Recall, a version of the listening span task, and Backwards Digit Recall, were also poorer in the SLI group but between-group differences were smaller, with smaller differences in the means and smaller effect sizes. Again this suggests a severe limitation in STM in the SLI group, combined with a milder deficit in WM. This pattern has been supported by other studies of STM / WM in SLI (Archibald & Gathercole, 2007).

With regard to the ALI children, the Group by Length interaction suggests that their NWR difficulties may have a different underlying origin, as suggested by Whitehouse et al. (2008). In particular the concave error profile in the SLI group indicates that STM limitations play an important role, and this claim is supported by the finding of poor performance on a range of STM assessments (DR, BDR, LS). However, further analyses investigating potential qualitative differences between the groups, with regard to errors involving difficulties with phonological structure, or phonemic encoding, failed to identify robust between-group differences. Another possibility presents itself, that the difference in performance is merely due to the severity of the deficit in the SLI group, not the influence of a separate factor in the ALI group. Unlike the Whitehouse et al. (2008) study, a straightforward comparison of error rates, this time calculated on a phoneme-by-phoneme basis indicated *more severe* difficulties in the SLI group, with a marginally significant effect ($p = .050$). This finding is particularly striking given that the participants with SLI were significantly older than the participants with ALI. The coding of errors at the phoneme level, as opposed to the word-level analysis in Whitehouse et al. (2008), may have enhanced the sensitivity of the NWR assessment used in the current study. The qualitative difference in the error profiles of the two groups may arise because the manifestation of an underlying factor is mediated by stimulus length in a *non-linear* fashion such that the greater the STM limitation, the larger the discrepancy in errors between the short and long stimuli. From this perspective, we might argue that the two groups have similar difficulties, which merely happen to be more severe in the individuals with SLI recruited for the current study, and therefore the phenotypic overlap hypothesis is preserved.

Ultimately, it is difficult to make any inferences about the phenotypic overlap between ASD and SLI purely on the basis of quantitative or qualitative differences in performance on a single task. However, the findings of the current study, and Whitehouse et al. (2008), can be viewed in the context of converging data suggesting qualitative differences between individuals with SLI and ALI on language tasks. For example, poor performance on past tense tasks in individuals with ALI group may be due to high rates of null responses, as opposed to morphological difficulties (Williams et al., 2008), and sentence repetition in adolescents with ASD is less sensitive to syntactic complexity than in adolescents with SLI (Riches, Loucas, Baird, Charman, & Simonoff, 2010). Overall these qualitative differences suggest that the claim for a phenotypic overlap between SLI and ALI may have been overstated. Nonetheless, more research is clearly needed in this area.

The qualitative analysis found an approximate 60:40 ratio between structure-changing and structure preserving errors, and an almost total lack of syllable omission or addition errors. This finding that errors predominantly preserved structure provides some support for the claims of Marton and Schwartz (2003). However, structure-changing errors were also clearly frequent. This suggests that the kind of syllable-altering changes identified by Gallon and colleagues (2006) may extend beyond the narrowly defined population of Grammatical SLI, and into the population of children with SLI whose diagnosis is based on omnibus assessments. The finding of high rates of both structure-changing and structure-preserving errors suggests the existence of two separate causal factors, a deficit affecting phonologically complex structures, as proposed by Gallon et al., and difficulties with simultaneous processing as proposed by Marton and Schwartz (2003). An interesting finding in the current data was that the ratio of structure-changing errors declined significantly as a function of word-length. Therefore C-for-C and V-for-V substitution errors, the structure-*preserving* errors, may be more closely associated with STM-load, further boosting the claim that such errors are related to attentional / memory processes (K. Marton & Schwartz, 2003), while structure changing errors reflect difficulties with phonological representation, and are therefore not linked to factors such as length.

Analysis of consonant substitutions identified a similar pattern to Shriberg et al. (2009), with a greater proportion of manner changes in the language-impaired groups, which may be indicative of auditory-perceptual encoding difficulties. However, between-group differences were non-significant and effect-sizes were small (0.07 compared to 0.73 in the Shriberg et al. study). Differences between studies may be due to differences in the size of the SLI group (13 versus 10), or age differences (mean 15;4 versus 4;8). Given that C-for-C substitutions are frequent during NWR, and are easy to code in terms of phonetic features, this kind of analysis has some potential for distinguishing between different clinical groups. However, data from the current study suggest that differences between clinical and non-clinical groups on this measure may not be large.

The use of an automatic algorithm, the Levenshtein Distance, may offer considerable advantages for data analysis. Firstly, it may be a more sensitive assessment of repetition. On a theoretical level it is clearly capable of distinguishing between children who make few phoneme errors per word, and children who make numerous phoneme errors. Such a pattern would not be detected by an all-or-none scoring scheme. This greater sensitivity may explain why differences between the ALI and SLI, i.e. a marginally significant effect of Group, and a significant Length x Group interaction, were evident only when the dependent variable was the phoneme-level LD. Equivalent analyses using the all-or-none scoring did not find significant between-group differences. One current NWR task (Dollaghan & Campbell,

1998) does in fact code on a phoneme-by-phoneme basis, and the LD may facilitate this scoring process. Furthermore, the use of an algorithm such as the LD facilitated qualitative analysis in that it allowed us to compare error rates for phonemic versus CV transcriptions, thereby enabling an analysis of the ratio of structure-preserving to structure-changing errors. Automated error analysis may lead to the development of new qualitative analyses which may be too complex and / or time-consuming to be performed by hand. In particular, the LD may prove useful in diagnostic assessments, given its greater sensitivity.

A strength of the current study is that by studying adolescents we reduced the role of confounding factors such as speech difficulties, typically present in young children, and therefore focused on core mechanisms involved in persistent NWR difficulties. However, this also complicates the task of comparing findings across studies, e.g. comparing the current study with Whitehouse et al. (1998). Further studies may wish to analyse the repetition performance of younger children using the techniques developed for the current study. While the qualitative analyses investigating structural changes and C-for-C substitutions did not yield significant differences, they nonetheless identified theoretically interesting patterns, and may prove to be useful methodologies in investigation of NWR difficulties in children.

Acknowledgments

The authors wish to thank Autism Speaks/The National Alliance for Autism Research for their generous funding; the parents/guardians and individuals who participated in the study; and Susie Chandler, Abigail Davison-Jenkins, Ann Ozsivadjian, and Vicky Slonims for their help in screening the participants with autism spectrum disorders (ASD). The first author wishes to thank Colin Bannard at the Max Planck Institute, Leipzig, Germany, for explaining the concept of the Levenshtein Distance; and Michael Gilleland of Merriam Park software for publishing a procedure to calculate the Levenshtein Distance on the internet, which was used during data analysis.

References

Adams, A., & Gathercole, S. E. (1996). Phonological working memory and spoken language development in young children. *The Quarterly Journal of Experimental Psychology*, *49*(1), 216-233.

Adams, A. M., & Gathercole, S. E. (1995). Phonological working memory and speech production in preschool children. *Journal of Speech and Hearing Research*, *38*, 403 - 414.

American Psychiatric Association. (2000). *Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition, Text Revision (DSM-IV-TR)*. Washington, DC.: American Psychiatric Association.

Archibald, L. M., & Gathercole, S. E. (2007). Nonword repetition in specific language impairment: more than a phonological short-term memory deficit. *Psychonomic bulletin & review*, *14*(5), 919.

Bailey, T. M., & Hahn, U. (2005). Phoneme similarity and confusability. *Journal of Memory and Language*, *52*(3), 339-362.

Baird, G., Simonoff, E., Pickles, A., Chandler, S., Loucas, T., Medlrum, D., & Charman, T. (2006). Prevalence of disorders of the autism spectrum in a population cohort of children in South Thames - the Special Needs and Autism Project (SNAP). *The Lancet*, *368*(9531), 210 - 215.

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

Ceponiene, R., Lepisto, T., Shestakova, A., Vanhala, R., Alku, P., Naatanen, R., & Yaguchi, K. (2003). Speech-sound-selective auditory impairment in children with autism: they can perceive but do not attend. *Proceedings of the national academy of sciences*, *100*(9), 5567 - 5572.

Conti-Ramsden, G., Botting, N., & Faragher, B. (2001). Psycholinguistic markers for Specific Language Impairment (SLI). *Journal of Child Psychology and Psychiatry*, *42*, 741-748.

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

Gathercole, S. E., & Baddeley, A. D. (1989). *Nonword Memory Test*. Centre for Working Memory and Learning, the University of York. Retrieved from http://www.york.ac.uk/res/wml/test of WM.html

Gathercole, S. E., & Baddeley, A. D. (1996). *The children's test of non-word repetition*. The Psychological Press.

Gathercole, S. E., & Pickering, S. (2001). *Working Memory Test Battery for Children*

(WMTB-C). Harcourt Assessment.

Gathercole, S. E., Tiffany, C., Briscoe, J., Thorn, A., & ALSPAC. (2005). Developmental consequences of poor phonological short-term memory function in childhood: a longitudinal study. *Journal of child Psychology and Psychiatry, 2005*, 598 - 611.

Gathercole, S. E., Willis, C., Emslie, H., & Baddeley, A. D. (1992). Phonological memory and vocabulary growth during the early school years: a longitudinal study. *Developmental Psychology, 28*, 887 - 898.

Kjelgaard, M. M., & Tager-Flusberg, H. (2001). An investigation of language impairment in autism: implications for genetic subgroups. *Language and Cognitive Processes, 16*(2 / 3), 287 - 308.

van der Lely, H. K. J., & Gallon, N. (2006). Exploring the impact of higher level linguistic representations on nonword repetition performance. *Applied Psycholinguistics, 27*(04), 591-594.

Levenshtein, V. I. (1966). Binary codes capable of correcting delections, insertions, and reversals. *Soviet Physics - Doklady, 10*, 707 - 710.

Lord, C., Risi, S., Lambrecht, L., Cook, E. H., Leventhal, B. L., DiLavore, P. C., Pickles, A., et al. (2000). The Autism Diagnostic Observation Schedule, Generic: A Standard Measure of Social and Communication Deficits Associated with the Spectrum of Autism. *Journal of Autism and Developmental Disorders, 30*(3), 205-223.

Lord, C., Rutter, M., & Couteur, A. (1994). Autism Diagnostic Interview-Revised: A revised version of a diagnostic interview for caregivers of individuals with possible pervasive developmental disorders. *Journal of Autism and Developmental Disorders, 24*(5), 659-

685.

Marton, K., & Schwartz, R. (2003). Working memory capacity and language processes in children with Specific Language Impairment. *Journal of Speech Language and Hearing Research*, *46*, 1138-1153.

Marton, K. (2006). Do nonword repetition errors in children with specific language impairment reflect a weakness in an unidentified skill specific to nonword repetition or a deficit in simultaneous processing? *Applied Psycholinguistics, 27*(04), 569-573.

Munson, B., Kurtz, B. A., & Winsor, 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*(5), 1033.

Roberts, J. A., Rice, M. L., & Tager-Flusberg, H. (2004). Tense marking in children with autism. *Applied Psycholinguistics*, *25*, 429 - 448.

Semel, E., Wiig, E. H., & Secord, W. (2000). *Clinical Evaluation of Language Fundamentals - Third Edition UK (CELF-3UK)*. Harcourt Assessment.

Shriberg, L. D., Lohmeier, H. L., Campbell, T. F., Dollaghan, C. A., Green, J. R., & Moore, C. A. (2009). A Nonword Repetition Task for Speakers With Misarticulations: The Syllable Repetition Task (SRT). *Journal of Speech, Language, and Hearing Research*, *52*(5), 1189.

Stothard, S. E., Snowling, M. J., Bishop, D. V. M., Chipchase, B. B., & Kaplan, C. A. (1998). Language-impaired preschoolers: A follow-up into adolescence. *Journal of Speech, Language and Hearing Research, 41*(2), 407.

Tager-Flusberg, H., & Joseph, R. (2003). Identifying neurocognitive phenotypes in autism. *Philosophical Transactions of the Royal Society, B., 358*, 303-314.

Vance, M., Stackhouse, J., & Wells, B. (2005). Speech-production skills in children aged 3-7 years. *International Journal of Language & Communication Disorders*, *40*(1), 29-48.

Wechsler, D. (1992). *Wechsler Intelligence Scale for Children - 3rd UK Edition (WISC-III UK)*. Pearson.

Whitehouse, A. J. O., Barry, J. G., & Bishop, D. V. M. (2008). Further defining the language impairment of autism: Is there a specific language impairment subtype? *Journal of Communication Disorders*, (4), 319 - 336.

Williams, D., Botting, N., & Boucher, J. (2008). Language in autism and specific language impairment: where are the links? *Psychological bulletin*, *134*(6), 944.

World Health Organisation. (1993). *The ICD-10 classification of mental and behavioural disorders: diagnostic criteria for research*. Geneva, Switzerland: World Health Organisation.

Table 1

	TD n = 17 (10 male)	SLI n = 13 (all male)	ALI n = 16 (all male)	Group comparisons
Age	172 (14y;4m) 4.19 168 - 179	184 (15y;4m) 7.26 173 - 199	176 (14y;8m) 5.77 168 - 184	
WISC Mean subtest SS	11.2 2.89 7 - 16.5	11.7 2.26 8.5 - 17.5	11.6 2.44 8 - 16	F(2, 43) = 0.169 p = 0.845 partial $\eta^2 = 0.008$
CELF Mean subtest SS	9.59 1.75 7.5 - 14	4.12 0.98 3 - 6	4.69 1.083 3 - 6.5	t(27) = -1.478 p = 0.151 d = 0.572
CD Raw	28.2 1.47 25 - 30	22.1 3.62 12 - 28	21.2 4.71 11 - 27	t(27) = 0.560 p = 0.580 d = 0.053
RS raw	63.5 6.84 49 - 75	34.2 8.18 24 - 54	38.4 11.3 18 - 58	t(27) = -1.122 p = 0.272 d = 0.045
DR raw	36.7 7.49 24 - 52	26.5 3.95 20 - 32	30.8 5.46 22 - 44	t(27) = -2.370 p = 0.025* d = 0.196
DR SS	107 19.9 74 - 143	78.6 10.4 61 - 93	90.9 15.2 68 - 129	
BDR Raw	18.6 5.93 12 - 30	11.2 3.31 6 - 17	14.2 4.71 7 - 20	t(27) = -1.960 p = 0.060 d = 0.190
BDR SS	101 14.9 84 - 129	80.3 8.73 68 - 96	88.6 12.6 68 - 104	
LR Raw	16.5 3.22 11 - 26	11.5 4.82 5 - 25	12.9 4.31 5 - 20	t(27) = -0.824 p = 0.417 d = 0.073
LR SS	107	80.6	88.7	

13.7	22.7	21.1
80 - 142	56 - 144	56 - 123

CD = Concepts and Directions, RS = Recalling Sentences, DR = Digit Recall, BDR = Backwards Digit Recall, LR = Listening Recall

Table 2

Mean LD per word for phonemic transcription by Length and Group

Means, standard deviations and ranges

	TD	SLI	ALI
All stimuli	0.22 0.14 *0.1 - 0.7*	0.69 0.39 *0.2 - 1.5*	0.45 0.23 *0.2 - 1.5*
2 syllables	0.16 0.08 *0.1 - 0.4*	0.28 0.21 *0.1 - 0.8*	0.26 0.12 *0 - 0.5*
3 syllables	0.21 0.18 *0 - 0.7*	0.37 0.18 *0.1 - 0.7*	0.30 0.18 *0.1 - 0.73*
4 syllables	0.22 0.18 *0.1 - 0.8*	0.90 0.55 *0.1 - 1.9*	0.53 0.23 *0.2 - 0.9*
5 syllables	0.28 0.23 *0 - 0.9*	1.28 0.87 *0.3 - 3.2*	0.76 0.63 *0.1 - 2.6*

Figure 1

Mean error rate by Group and Syllable Length

Dark bars show LD for phonemic coding. Light bars show LD for CV coding.
Whiskers show standard errors