To appear in *Journal of Speech, Language, and Hearing Research*

The Preschool Repetition Test: An evaluation of performance in typically developing and clinically referred children

Shula Chiat

e-mail: shula.chiat.1@city.ac.uk

Penny Roy

e-mail: p.j.roy@city.ac.uk

Address for correspondence:

Shula Chiat

Department of Language and Communication Science

City University

Northampton Square

London EC1V 0HB

Abstract

Purpose: To determine the psychometric properties of the Preschool Repetition Test (Roy & Chiat, 2004); to establish the range of performance in typically developing children and variables affecting this; and to compare the performance of clinically referred children.

Method: The PSRep Test comprises 18 words and 18 phonologically matched nonwords systematically varied for length and prosodic structure. This test was administered to a ‘typical’ sample of children aged 2;0–4;0 (n=315) and a ‘clinic’ sample of children aged 2;6-4;0 (n=168), together with language assessments.

Results: Performance in the typical sample was independent of gender and SES, but was affected by age, item length, and prosodic structure, and was moderately correlated with receptive vocabulary. Performance in the clinic sample was significantly poorer, but revealed similar effects of length and prosody, and similar relations to language measures overall, with some notable exceptions. Test-retest and interrater reliability were high.

Conclusions: The PSRep Test is a viable and informative test. It differentiates within and between ‘typical’ and ‘clinic’ samples of children, and reveals some unusual profiles within the clinic sample. These findings lay the foundations for a follow-up study of the clinic sample to investigate the predictive value of the test.

Interest in nonword repetition as an assessment tool has developed considerably over the last decade. This interest stems from Gathercole and Baddeley's (1989) finding that nonword repetition performance correlated with performance on a variety of language measures in typically developing children aged 4 to 5, and subsequent corroboration in further studies covering a wider age range (Adams & Gathercole, 2000; Gathercole, Willis, Emslie, & Baddeley, 1992; Gathercole, Willis, Baddeley, & Emslie, 1994). These findings provided the impetus for investigating nonword repetition in atypically developing children. Would nonword repetition performance be indicative of language skills – and language deficits - in these children? Findings are consistent in showing differences in nonword repetition performance of typically and atypically developing children at a group level (Bishop, North, & Donlan, 1996; Botting & Conti-Ramsden, 2001; Conti-Ramsden, Botting, & Faragher, 2001; Dollaghan & Campbell, 1998; Edwards & Lahey, 1998; Ellis Weismer, Tomblin, Zhang, Buckwalter, Chynoweth, & Jones, 2000; Gathercole & Baddeley, 1990; Gray, 2003; Marton & Schwartz, 2003; Montgomery, 2004; Norbury, Bishop, & Briscoe, 2002). Some studies have found that nonword repetition is a relatively reliable indicator of SLI (Bishop, North, & Donlan, 1996; Conti-Ramsden, Botting, & Faragher, 2001; Dollaghan & Campbell, 1998), and even picks out children previously diagnosed with SLI but whose language difficulties appear to have resolved (Bishop, North, & Donlan, 1996; Conti-Ramsden, Botting, & Faragher, 2001). For these reasons, Bishop, North and Donlan and Conti-Ramsden, Botting and Faragher have proposed nonword repetition as a possible marker for SLI. Edwards and Lahey (1998), on the other hand, found overlap between the nonword repetition scores of children with and without SLI. While some studies have used the Children's Nonword Repetition Test (Gathercole & Baddeley, 1996), some have used Dollaghan and Campbell's (1998) Nonword Repetition Test, and others (such as Edwards and Lahey, 1998) have used their own set of stimuli. Differences in the characteristics of the test items, along with other methodological differences such as age of participants and particular criteria for identification of SLI, may account for differences in the extent to which nonword repetition is found to discriminate groups.

While the precise status of nonword repetition as a marker of SLI remains uncertain, well substantiated evidence of its relationship to language skills indicates that it is a potentially valuable assessment tool. It is quick and easy to administer, and poor performance on it points to likely problems, which can then be investigated further.

These findings and observations led to our hypothesis (Roy & Chiat, 2004; see also Chiat, 2001) that evidence of repetition skills may be especially useful in the assessment of very young children. Methodologically, the task is undemanding, eliciting a behaviour that children produce spontaneously. It is performed relatively automatically, in the sense that it requires on-line speech input, temporary storage, and output processes; in contrast to tasks such as picture description, picture pointing, and same-different or yes-no judgements, it does not require interpretation of, or decisions about, linguistic stimuli.

Though a few studies have assessed children under 4 years (Adams & Gathercole, 1995; Gathercole & Adams, 1993; Vance, Stackhouse, & Wells, 2005), nonword repetition tests to date have been standardised on older children (ages 4 to 8 years), and on the whole are too demanding for younger children. The Preschool Repetition Test (Roy & Chiat, 2004) was designed to be realistic for this age group, quick to administer, maximally informative, and in line with developmental capabilities. The test includes both words and nonwords, enabling us to investigate the effects of familiarity on repetition performance. Both real word and nonword targets are prosodically controlled, allowing us to investigate the effects of prosodic structure on repetition performance. The test yields two measures of performance: accuracy (number of items repeated correctly), and rate of syllable loss.

Roy and Chiat (2004) reported the results of the first trial of the test with a sample of 66 typically developing children aged 2-4 years. Consistent with findings in studies on older children, children's performance on the test was sensitive to age, but not to gender, or an estimate of socioeconomic status. Consistent with our hypotheses, children's performance was sensitive to word status, with higher scores on words than nonwords, indicating that children typically benefited from lexical familiarity. Unsurprisingly, responses were sensitive to item length. They were also sensitive to prosodic structure, with virtually no loss of stressed syllables, and with loss of unstressed syllables varying according to their position relative to stress: loss of prestress syllables far exceeded loss of poststress syllables. In line with many, though not all, previous studies, repetition performance was moderately correlated with performance on the British Picture Vocabulary Scales (Dunn, Dunn, Whetton, & Burley, 1997), a test of receptive vocabulary.

On the basis of these findings, we concluded that our word and nonword repetition task had potential as a clinically informative assessment tool. The present paper further examines this potential. It replicates the 2004 study with a larger sample of 315 children spanning different social and geographical backgrounds. This enabled us to verify and extend our findings on the effects of gender and age of participants, and lexical status and prosodic structure of items. Replication of our 2004 study also enabled us to evaluate the psychometric validity of the test by including measures of test-retest as well as interrater reliability. Test-retest reliability has previously been investigated by Gray (2003), who administered a nonword repetition test to children with SLI and with normal language (age 4;0-5;11) on two consecutive days. She found significant improvement from first to second administration for both groups, with greater improvement in the SLI group; on the other hand, correlations between test times were significant for the SLI group but not for the normal language group. Despite the observed changes, Gray reports that test sensitivity and specificity remained high.

The replication of our earlier study, reported below, provides the data from typically developing children needed for the second study reported in this paper, in which we administered the test to 168 clinically referred children. The goal was to find out whether and how the performance of the clinic sample differed from that of the typically developing sample.

We predict that:

i. Accuracy scores in the clinic sample will be significantly lower than in the typical sample.

ii. Rate of syllable loss in the clinic sample will be significantly higher than in the typical sample.

On the other hand, we predict that, as with the typical sample:

iii. Performance, as measured by accuracy scores and rate of syllable loss, will improve with age.

iv. Words will show an advantage over nonwords for the clinic sample as a whole.

v. The range of advantage for words over nonwords will be greater than the range of advantage for nonwords over words.

vi. Syllable loss will be sensitive to length and to prosodic structure, and will show the same pattern of difficulty observed in the typical sample, with loss of prestress syllables > poststress syllables > stressed syllables.

Study 1: The typical sample

Method

Participants

A sample of 333 children was recruited from nurseries and playgroups whose heads were willing to distribute invitations to parents. These facilities were spread between urban, suburban and rural areas of England. All children in the targeted age range (2;0-4;0) with English as a first language and no history of hearing loss, no referral to speech and language therapy services, and whose parents gave consent, were included in the study. Information about socioeconomic status, geographical area and language use was obtained through a short questionnaire administered verbally to heads of nurseries. SES was broadly classified as upper, middle or lower class. This terminology is widely used in the UK. Judgements of class are typically based on an amalgam of parental background, occupation, and home environment. Complete information was available for just over three-quarters of the children. The sample comprised 48% boys and 52% girls. Just under a third were in state- or charity-funded nurseries (30%), two-thirds (64%) in privately-run nurseries, and the rest (6%) in nurseries with a sliding scale of fees. According to the reports of nursery staff, two-thirds (66%) were middle class, just under a quarter (23%) were lower class, and a tenth (11%) were upper class. For the majority of children (87%), only English was spoken at home, with 13% exposed to additional languages.

Of the children recruited, 6% refused to participate in the repetition task (see below), giving a final sample of 315 children, with a mean age of 35.9 months, SD=6.08. Since the refusers complied with a nonverbal test of receptive vocabulary, and their performance did not differ from the rest of the sample on this test [$M_{complied}$ = 29.62, $SE_{complied}$ = .69; $M_{refused}$ = 31.82, $SE_{refused}$ = 2.91; F(1,329) = .54, *p* = .46, including age as a covariate], it is likely that their refusal on the repetition task reflected verbal reticence rather than inability.

Test items

The PS Rep Test consists of 36 test items, comprising a block of 18 words followed by a block of 18 nonwords. Words and nonwords are equally divided in length between one-, two- and three-syllable items. They are of identical prosodic structure, with patterns of strong syllables (with primary or secondary stress) and weak (unstressed) syllables systematically manipulated across two- and three-syllable items. They are phonologically matched in that nonwords were created by altering the vowel of one-syllable words, e.g. lamb → /lɒm/, and transposing two or three of the consonants in two- and three-syllable words, e.g. *banana* → /nəˈnɑbə/, *holiday* → /ˈlɒdɪˌheɪ/. The test includes 4 practice items comprising 2 words and 2 phonologically matched nonwords. These were used to introduce the test words and nonwords, and were not scored. (See Appendix for the full set of items).

Scoring

The PSRep Test yields two measures of performance: number of whole items correct, and number of syllables lost. Responses were scored as correct if they contained all phonemic segments of the target in the correct order, with no additional phonemes. Allowances were made for local/regional dialect, and for substitutions that were either consistent or phonetic variants (e.g. dentals for alveolar targets). This scoring by whole items correct is similar to the scoring used in the CNRep Test (Gathercole et al., 1994). It was favoured over scoring by percent phonemes correct used in some tests, for example Dollagan and Campbell's (1998), on the grounds that whole item scoring is equally differentiating and much quicker to calculate (Gray, 2003; Roy & Chiat, 2004). Responses were also scored for syllable loss (for rationale, see Roy & Chiat, 2004). Loss of a syllable was recorded where

(a) a vowel was omitted with/without adjacent consonants, e.g. *holiday* → /ˈhɒˌdeɪ/ or /ˈhɒlˌdeɪ/ (with two exceptions, since /blun/ and /plis/ were considered acceptable realisations of *balloon* and *police*)

(b) two syllables were coalesced, combining the consonant from one with the vowel from the other, e.g. *balloon* → /bun/.

Procedure

Each child was seen individually, in their nursery or playgroup. Order of presentation was the same for all children. First, a test of receptive vocabulary, the British Picture Vocabulary Scales (BPVS; Dunn et al., 1997) was administered. The words and nonwords of the PSRep Test were then presented, starting with the practice items, which were used to familiarise the child with the task before scoring started. The child was not required to respond or provide a correct response to the practice items in order to proceed.

The PSRep Test was presented live. Most studies of nonword repetition with older children use recorded stimuli, one of the key reasons being that this eliminates any visual cues. In a study involving 3-year-olds, Adams and Gathercole (1995) presented stimuli live, but with the experimenter's mouth hidden from view. Arguments can be made for both methods, and careful consideration was given to these bearing in mind the aims of the PSRep Test. The test was designed to investigate very young children's ability to repeat words and nonwords under conditions that optimised the chances of eliciting a response. Since it was not designed as a test of auditory short-term memory, availability and use of visual cues – which normally accompany speech input – was not a problem for our purposes. The ultimate aim of the test was to serve as a clinical tool. Very young children and clinically referred children with poor attention are more likely to produce responses if the tester actively engages with them. Fisher, Hunt, Chambers, and Church (2001) administered 32 monosyllabic nonwords to 53 children aged 28-31 months. Fewer than half those tested (24 children) provided sufficient responses to be included in the study. In further studies with 3-year-olds, a third of the children failed to produce sufficient responses. High nonresponse rates in our samples would reduce their representativeness, and would limit the usefulness of the test in clinical assessment.

On the other hand, live presentation has some disadvantages. Use of recorded stimuli ensures uniformity of input, eliminating variations in rate, pitch, volume and other phonetic and auditory features of input which may occur when the tester delivers the stimuli, and which may enhance or depress children's performance. Systematic comparison of live versus recorded stimuli is clearly needed to evaluate the possible effects of such variations on rates and accuracy of response, and the comparability and validity of the two methods. In the absence of relevant evidence, the possible effects of live presentation must be considered in interpreting our findings.

The PSRep Test was presented using a puppet with a movable mouth. The child was introduced to the puppet and asked to copy some words that the puppet was going to say. The nonwords were presented in the same way except that, in this case, stimuli were described as “silly puppet words” rather than words. Frequent verbal praise and stickers were used to encourage maximum levels of participation from the children. If a child failed to respond to an item, up to two further opportunities were given.

Each child's responses were audiorecorded, transcribed and scored by research assistants who were trained and highly proficient in phonetic transcription. Responses of a random subsample of 60 children were transcribed and scored independently by a second research assistant in order to evaluate interrater reliability. In order to evaluate test-retest reliability of the PSRep Test, a random subsample of 44 children were retested within two weeks.

Results

For all results, Bonferroni corrections for multiple comparisons were applied and Games Howell and Gabriel procedures adopted to deal with unequal sample sizes and variances. In cases where the exact probabilities varied, the most conservative estimate is reported. An alpha level of .05 was taken for all statistical tests, and *p* values were two-tailed. Where appropriate, effect sizes (η^2) are reported.

Reliability

Interrater agreement. For the subsample of 60 children, the level of agreement between first and blind raters' scores for the total set of 36 items was high and of an acceptable level, with an intraclass correlation of $\alpha = 0.93$. Agreement for total syllable loss was also high, with Cohen's Kappa of $\kappa = .92$.

Internal consistency. Internal consistency of the total set of 36 items was calculated for the sample excluding refusers (n = 315), yielding a coefficient α of 0.92. Removal of any single item in the scale made virtually no impact on the resultant α values.

Test-retest reliability. Of the 44 children in the retest subsample, 3 refused the second round of testing. For the remaining 41 children, the intraclass correlation was high (α=0.93). The difference in total scores between the first and second round of testing was not significant ($mean_{difference}$ = -.68, SD = 3.11, t(40) = -1.41, p = .17).

Concurrent validity. Partial correlations between total items correct on the PSRep Test and on the test of receptive vocabulary (BPVS) were calculated. Raw scores rather than standard scores on the BPVS were used, as some children were younger than 3 years -- the starting age for standard scores on this test (this appeared to be justified: even the youngest children in the sample performed above chance level). Partial correlations were in line with previously reported findings (Roy & Chiat, 2004): moderate but significant including age as a covariate (r = 0.3, p < 0.001). Likewise, dividing the group into higher and lower vocabulary scorers (BPVS raw scores above versus at/below the median for their age group) yielded significant differences for word, nonword and total word/nonword scores with age partialled out [$F_{word}(1,311) = 5.76$, $p = .017$, $\eta^2 = .018$; $F_{nonword}(1,311) = 13.75$, $p < .001$, $\eta^2 = .04$; $F_{word/nonword}(1,311) = 10.76$, $p = .001$, $\eta^2 = .03$].

Nonresponses

Once children started the test, nonresponses were rare: 64% of the entire sample attempted every single item and a further 13% missed only one item. Only 8% produced 6 or more nonresponses. As pointed out above, 6% refused altogether and were excluded from subsequent analysis. A related t test showed that word status affected children's nonresponse rate: the rate of nonresponse for words was significantly lower than the rate of nonresponse for nonwords ($mean_{word} = .91$, $SD_{word} = 2.88$; $mean_{nonword} = 1.28$, $SD_{nonword} = 3.91$; t(315) = -2.94, p = .004).

Participant variables

Univariate analyses were carried out to investigate the effects of gender (2 levels); type of nursery (3 levels: state/charity, private, mixed); socioeconomic status

(SES; 3 levels: upper, middle, lower); and language(s) spoken at home (2 levels: English only, one or more other languages). As four separate analyses were run, an adjusted alpha level of .0125 was adopted. Since repetition performance was previously found to be age-related (Roy & Chiat, 2004), confirmed by a significant correlation between age and performance in the present study ($r = .44$, $p < .001$), these univariate analyses included age as a covariate. None of the demographic factors were found to have a significant effect [$F_{SES}(2,234) = .3$, $p = .7$; $F_{gender}(1,312) = .7$, $p = .4$; $F_{language}(1,235) = .16$, $p = .7$; $F_{nursery}(2,311) = 2.56$, $p = .08$, post hoc comparisons: $p_{state/private} = 1$]. These factors were therefore excluded from subsequent analyses.

On the other hand, SES was found to have a significant effect on receptive vocabulary scores [$M_{upper} = 36.09$, $SE_{upper} = 2.37$; $M_{middle} = 34.48$, $SE_{middle} = 0.97$; $M_{lower} = 26.82$, $SE_{lower} = 1.62$; $F(2, 233) = 9.24$, $p < .001$].

Overall scores

Overall scores for the 36 items ranged from 1 to 36. Across the entire sample of children, over 80% received scores of 20-36. These results are highly consistent with our previous study, where the sample included a smaller proportion of children under 2;6, and 90% of the total sample scored in the 20-36 range.

Table 1 gives a breakdown of overall scores by age, lexical status, and item length. It should be noted that nonresponses were treated as ‘incorrect’ when totalling number of items correct.

Table 1 about here

Age

A univariate analysis taking 4 levels of age (2;0<2;6, 2;6<3;0, 3;0<3;6, 3;6-4;0) revealed that age had a significant effect on repetition [$F(3,315) = 27.16$, $p = <.001$,

η^2=.21]. Post hoc comparisons for age revealed no significant difference between the two older age groups ($M_{3;0<3;6}$ = 28.5, $SD_{3;0<3;6}$ = 5.16; $M_{3;6-4;0}$ = 28.8, $SD_{3;6-4;0}$ = 6.24; *p* = 1). These two age groups were therefore combined into a single group aged 3;0-4;0. A mixed ANOVA was carried out with two within-factors (word status: two levels, item length: three levels), and one between-factor (age: 3 levels). As expected age with the three-level age factor was significant [$F(2,312) = 40.48$, $p < .001$, η^2=.21]. Post hoc comparisons confirmed that differences between all three age groups were significant ($p < .001$ for all comparisons except the two younger groups where $p = .03$). Boxplots for the typical and clinic samples (see Figure 1 below) illustrate differences in scores and ranges across the age groups, and the presence of outliers (most of whom were in the oldest group)[1].

Item variables

Mean scores and standard deviations according to word status and item length are shown in Table 1. As in our previous study, both factors were found to have significant effects on children's repetition performance. A significant effect of word status was found, with an advantage of words over nonwords [$F(1,312) = 138.31$, $p < .001$, $\eta^2 = .31$]. In line with this, the majority of children repeated words more accurately than nonwords. This difference could not be explained by the difference in nonresponse rates for words and nonwords reported above. Comparable results were obtained when the analysis was repeated for only those children who attempted every item [$F(1,211) = 65.74$, $p < .001$, $\eta^2 = .24$]. Taking the children who attempted at least 31 items on the test (91% of the sample), only 13.6% scored higher on nonwords, compared with 72.5% on words (with the rest showing no difference). Furthermore, the advantage for words ranged from 1 to 11, while the advantage for nonwords was never greater than 4.

The ANOVA also revealed a significant effect of item length [$F(2,624) = 338.16$, $p < .001$, $\eta^2 = .52$]. All post hoc comparisons were significant ($p < .001$): accuracy of repetition was greatest for monosyllabic items and lowest for 3 syllable items (Table 1). Significant two-way interaction effects were found for word status by length [$F(2,624) = 7.82$, $p < .001$, $\eta^2 = .02$] and three-way interaction effects for word status by length by age [$F(4,624) = 2.66$, $p = .032$, $\eta^2 = .02$]. However, effect sizes in both cases were small and post hoc comparisons were not pursued.

Syllable Loss and Prosodic Structure

As in our previous study, loss of whole syllables was relatively rare. Just under a quarter of the children (21.6%) never omitted or coalesced syllables. Of the remaining children, the majority lost no more than four syllables across their entire set of responses; just under a quarter (22.5%) omitted five or more syllables, and the maximum number of syllable omissions was 16 (two children).

Table 2 shows the mean and range of syllable loss according to prosodic position of the syllable, age, and item variables of lexical status and length. These are presented as a percentage of the number of target syllables children attempted, i.e. excluding nonresponses.

Table 2 about here

Given the floor effects, skewed distribution, and unequal variances of the data, a parametric analysis of these errors was not possible, so chi-square tests were used. Analysis of total syllable loss reveals striking effects of prosodic structure ($\chi^2 = 1704.2$, df = 6, $p < .001$). Stressed syllables were preserved almost without exception, regardless of word length. Secondary stressed syllables, which occurred only in three-syllable items, were relatively robust. The fate of unstressed syllables depended on their position in the prosodic structure. Those which occurred poststress, following a strong syllable in a trochaic structure, were relatively robust: in two-syllable items, they were preserved as well as stressed syllables, but they were more vulnerable in three-syllable items. Unstressed syllables which occurred prestress, outside a trochaic structure, were by far the most vulnerable. Compared with poststress syllables, they were about 25 times more liable to omission in two-syllable items, and about 3 times more liable in three-syllable items. Indeed, prestress syllables in two-syllable items were more vulnerable than poststress syllables in longer, three-syllable items (17.84% versus 7.87%). The profile of vulnerability is illustrated in Figure 2 below (which shows both typical and clinic samples). The rate of syllable loss changed with age ($\chi^2 = 247.9$, df = 2, $p < .001$: see Table 2). However, the relative vulnerability of different prosodic positions did not change with age.

As is also evident in Table 2, rate of syllable loss was not affected by lexical status ($\chi^2 = .159$, df = 1, $p = .69$), but was affected by item length ($\chi^2 = 36.7$, df = 1, $p < .001$), with significantly more loss in three-syllable than two-syllable items.

Discussion

This more extensive administration of the PSRep Test confirms and extends the findings of our smaller scale 2004 study.

First, the test is psychometrically robust, achieving high levels of interrater reliability, internal consistency, and test-retest reliability. Our finding of no difference between test-retest results contrasts with Gray's (2003) finding of significant improvement. This is most likely due to the longer interval between tests (1-2 weeks, compared with 1 day in Gray's study), but may also reflect age differences. Partial correlations with a test of receptive vocabulary, the BPVS, demonstrate concurrent validity of our test. They are also in line with partial correlations between nonword repetition and receptive vocabulary reported in previous studies (e.g. r=.34 in Gathercole and Adams' (1993) study of 3 year olds). Analysis in terms of levels of vocabulary scores confirmed that higher scorers had better repetition performance than lower scorers.

Second, in line with findings in our 2004 study and other studies (notably Ellis Weismer et al, 2000; Burt, Holm, & Dodd, 1999), gender**,** nursery type and SES did not influence performance, nor did number of languages spoken at home. In contrast to our findings on the PSRep Test, and in line with other studies (e.g. Hart & Risley, 1995), SES had a significant effect on receptive vocabulary performance. However, our measure of SES for the typical sample was a broad categorization. It is possible that wider sampling and finer-grained measures in future research may identify class differences in repetition skills at this age.

Third, the test yields a clear profile of development. At 2;0<2;6, a notable proportion of children (11%) were not compliant. Of those who did comply, over half obtained scores of 20 or above. By 2;6<3.0, non-compliance was rare, and performance was skewed towards the upper range, with over two-thirds scoring 20 or above. From 3;0, differences between children were levelling out, as evidenced by the lack of age effects between the age bands 3;0<3;6 and 3;6-4.0. Half the children in the year band 3;0-4;0 scored 30 or above, and 95% scored at least 20. The rate of syllable loss reduced with age, but the pattern of loss was consistent across age groups: prestress syllables were far more vulnerable than stressed or poststress syllables, regardless of item length.

As in our previous study, length significantly affected scores at all ages, with an advantage for shorter over longer items. This finding is consistent with other studies, which report length effects on different nonword repetition tests and at different ages (e.g. Dollaghan & Campbell, 1998; Ellis Weismer et al., 2000; Gathercole, Willis, Baddeley, & Emslie, 1994; Gray, 2003).

Three effects of lexical status were observed. Words were more likely to be attempted than nonwords, and they were repeated signficantly more accurately than nonwords. While associations with vocabulary were found for both words and nonwords, the effect size – although small in both cases - was stronger for nonwords. Other studies have investigated factors that affect repetition of words and nonwords, and found that real word repetition is influenced by neighborhood density, while nonword repetition is influenced by phonotactic probability (Edwards, Beckman, & Munson, 2004; Munson, Edwards, & Beckman, 2005; Munson, Swenson, & Manthei, 2005). Furthermore, phonotactic effects were found to correlate with vocabulary size. The stimuli in our test were designed to investigate length and prosodic structure of words and nonwords rather than neighborhood density or phonotactic probability. As our stimuli were not controlled for these factors, it is possible that they may have influenced children's repetition, and the relation between their repetition and vocabulary. Further investigations systematically manipulating these factors may throw more light on the extent and nature of the word-nonword difference and partial correlations with vocabulary observed in our study.

In conclusion, the profile of performance of the typical sample indicates that children rapidly develop the skills required for repetition of items up to 3 syllables, in which syllable structure is kept simple. Consequently, the time window for application of our test is narrow. It is most differentiating in the 2;6<3;0 age band, where response range was wide and compliance level reasonable; it is unlikely to be informative about children under 2 or over 4 who are developing normally. This is as we would hope, since our purpose in designing a preschool repetition test was to tap language-related skills that are robust in typically developing children, but may not be in children with language impairment. Comparing the performance of clinically referred children will reveal whether they show significant difficulties as a group. Equally importantly, given the inevitable heterogeneity of clinical referrals, we will discover whether some children in this group obtain scores for accuracy and syllable loss that fall within the typical range, and whether some show profiles of performance not seen in the typically developing children.

Study 2: The clinic sample

Method

Participants

Participants in the clinic sample were recruited from 4 inner London and 3 outer London Primary Health Care Trusts. Invitations to participate were sent to parents of children who were referred for speech and language therapy assessment and met the following criteria:

- aged 2;6-3;6
- reason for referral was concern about language development (not speech)
- no report of congenital problems, hearing loss, oro-motor difficulties, and no diagnosis of autism
- nonverbal ability within 2 SDs of the mean (≥70).

All children whose parents gave consent were included in the study. Overall, 209 children were recruited. Of these, 28 fell below the criterion for nonverbal ability and were excluded from the current analysis, yielding a sample of 181. A small proportion were over the targeted age limit, and were included in a separate 3;6-4;0 age group (n = 42). Of the sample of 181 children, 7% refused to participate in the repetition task, giving a final sample of 168 children, with a mean age of 37.6 months, SD = 4.84 (1.7 months older on average than the typical sample). This final sample had a mean nonverbal IQ of 91.2, SD = 12.1. The majority (61%) had nonverbal ability scores in the normal range (between 85 and 115); a small proportion (5%) had scores above average, with 34% below average. Turning to receptive and expressive language, mean scores for the sample were 87.8 and 84.8, SD = 15.6 and 13.6, respectively. Half the auditory scores (51%) and a third of the expressive scores (35%) were in the normal range; a small proportion (4% in both cases) had above-average scores, with 45% and 61% below average.

Three-quarters of the sample were boys. English was the first language for almost all children (96%), and the only language for three-quarters of the sample. The majority (71%) were white, 14% black, and the remainder described as Asian or of mixed heritage. Direct initial interview with the parent(s) of children in the clinic sample provided more detailed background information related to SES indices than was available for the typical sample. SES indices of education level and income levels in the clinic sample showed a wide distribution: a third of the sample had incomes below the national average, with a further third above. A third of the primary carers were educated to graduate level or above; just over a quarter (26%) left school at 16 or younger; 14% had left school with minimal educational qualifications or none. The most comparable measure in the clinic sample data to the information available for the typical sample is father’s occupation level. In order to make comparisons, the five way classification of father’s occupation was collapsed to form three categories that broadly correspond to the lower, middle and upper class categories adopted in the typical sample. Adopting this measure and excluding fathers who were described as students or househusbands in homes where the father was present, 28% were categorised as lower class, 60% as middle class and 12% as upper class (compared with 23%, 66% and 11% respectively in the typical sample).

Although criteria for referral to the study excluded children whose referral report made any reference to hearing difficulties, 23% of parents responded positively to a question on the parental interview asking whether the child had ‘any diagnosed or suspected hearing loss or glue ear’. However, almost all specified this as intermittent episodes of ‘glue ear’; only one mentioned the possibility of central hearing loss, but this was not confirmed, and the child’s hearing was considered normal at the time of the interview. The marginal nature of these problems was confirmed by the finding that hearing status was not related to repetition performance as measured by total PSRep score [$F(2,163) = .4$, $p = .67$ with the effect of age partialled out]. Likewise, there was no difference in repetition performance between children with English only and those exposed to more than one language [$F(1,165) = 1.3$, $p = .26$]. It was therefore not necessary to take these factors into account in subsequent analyses.

Procedure

Most children were seen on two occasions at home, with a parent present. A small number were assessed in their nursery or preschool language unit. Assessments were carried out by research assistants who were qualified speech and language therapists or linguists with relevant phonetic training and experience of assessing very young children. In most cases, two research assistants were present. The PSRep Test was administered as part of a battery of novel preschool assessments, together with standardised assessments of nonverbal ability (the British Ability Scales II: Elliot, 1996) and of auditory and expressive language (Pre-school Language Scale-3 (UK), Boucher & Lewis, 1998). In addition, parents were asked to complete the UK short version of the MacArthur Communication Development Inventory (MCDI-UKSF, Dale et al., 2000), a measure of early vocabulary development. Tests were carried out in the same order in each of the two sessions, but at a pace determined by the child. However, each test was administered as a whole, i.e. within one session. If children refused to participate, became agitated, or were clearly unable to comply with an assessment, testing was discontinued.

Procedure for test administration was as described above for the typically developing sample. Children's responses were transcribed online and videorecorded for subsequent checking. Responses for 41 children were scored independently by two researchers online, allowing for an evaluation of interrater reliability.

Results

Reliability

Interrater agreement. For the subsample of 41 children, the level of agreement between first and blind raters' scores for the total set of 36 items was high and of an acceptable level, with an intraclass correlation of $\alpha = .98$. Agreement for total syllable loss was also high, with Cohen's Kappa of $\kappa = .96$.

Relations with language measures

Partial Correlational analysis again revealed relations between performance on the PSRep Test and on assessments of language. With age partialled out, significant correlations were found between repetition scores and parent reported vocabulary (MCDI-UKSF) ($r = .52$, $p < .001$), and between repetition scores and both the auditory PLS ($r = .36$, $p < .001$) and the expressive PLS ($r = .43$, $p < .001$). Partialling out nonverbal ability scores as well as age made very little difference to levels of correlation ($r = .49$, .3, .38 respectively). Likewise, division into high and low vocabulary scorers (MCDI-UK_{SF} raw scores above versus at/below the median by age group) yielded significant differences for word, nonword and total word/nonword scores with age partialled out [$F_{word}(1,159) = 24.69$, $p < .001$, $\eta^2 = .13$; $F_{nonword}(1,159) = 19.45$, $p < .001$, $\eta^2 = .11$; $F_{word/nonword}(1,159) = 23.93$, $p < .001$, $\eta^2 = .11$].

Despite these associations, a small proportion of children showed a marked dissociation between repetition and language measures. Of the 81 children with very low repetition scores (<2 SDs for their age group), 10 achieved scores at or above the mean on the auditory PLS, and 5 on the expressive PLS. Conversely, of the 18 children with repetition scores at or above the mean for their age, 3 had low standard scores (<1SD) on the auditory PLS and 8 on the expressive PLS.

Nonresponses

The rate of total refusal was considerable in the youngest clinic group (13%), as it was in the youngest typically developing group (11%), bearing in mind that the typical sample starts at age 2;0 and the clinic sample only at 2;6. The youngest age group in the clinic sample also showed a notable rate of nonresponse: only 57% of the group produced fewer than 6 nonresponses (as did 67% of the youngest typically developing group). In the two older clinic age groups, rate of nonresponse was low: 84% aged 3;0<3;6 and 94% aged 3;6-4;0 produced fewer than 6 nonresponses. A related t test showed that, as was the case for the typical sample, word status affected children’s nonresponse rate: the rate of nonresponse for words was significantly lower than the rate of nonresponse for nonwords ($mean_{word} = 1.73$, $SD_{word} = 4.24$; $mean_{nonword} = 2.56$, $SD_{nonword} = 5.45$; $t(167) = -4.16$, $p < .001$).

Participant variables

Univariate analyses were carried out to investigate the effects of gender (2 levels); father’s occupation (3 levels); income (2 levels), and education (3 levels).

With an adjusted alpha level of .01 taken to allow for the multiple comparisons, and including age as a covariate, none of the demographic factors were found to have a significant effect on repetition performance ($p_{gender} = .66$; $p_{occupation} = .05$; $p_{income} = .06$; $p_{education} = .24$). On the other hand income level was found to have a significant effect on language as measured by the receptive PLS [$F_{income}(1,159) = 19.9$, $p < .001$, $\eta^2 = .11$]. The other factors fell short of significance ($p_{occupation} = .02$; $p_{education} = .08$; $p_{gender} = .8$).

Overall scores

Table 3 gives a breakdown of scores by age, lexical status, and item length. It should be noted that nonresponses were included in the overall score, where they were not differentiated from incorrect responses.

Table 3 about here

As for the typically developing children, a mixed ANOVA[2] was carried out with two within-factors (word status: two levels, item length: three levels), and one between-factor (age: 3 levels). As predicted, age with the three-way age divide was significant [$F(2,165) = 21.99$, $p < .001$, $\eta^2 = .21$]. In contrast to the typical sample, however, the post hoc comparisons showed that the difference between the two older groups in the clinic sample was significant ($p = .004$). Comparing these scores with the scores for the typical sample (Table 1), it is evident that the performance of the clinic sample on the PSRep Test was extremely weak: the means for each age group are 1.6-2 standard deviations below the means for their typically developing peers. The difference between the groups' performance is illustrated by the boxplots in Figure 1. It is striking that the median and range of the oldest clinically referred group (age 3;6-4;0) are closest to those of the youngest typically developing group who were on average 18 months younger (age 2;0<2;6).

Figure 1 about here

Statistical comparison confirmed these observations [$F(5,477) = 74.69$, $p < .001$, $\eta^2 = .44$]. Scores in the 2-6<3.0 and 3:0<3:6 were significantly below those of typically developing children in the same age bands ($p < .001$). Even more strikingly, they were significantly below the scores of typically developing children in the 2;0<2;6 band who were considerably younger ($p_{CG:2;6<3;0} < .001$, $p_{CG:3;0<3;6} = .004$) . In contrast post hoc comparisons showed that the oldest band in the clinic sample, aged 3;6-4;0, differed significantly from their age matched peers ($p < .001$) but they did not differ significantly from the two younger age bands in the typical sample. According to the non-parametric Mann-Whitney U test, the oldest clinic sample (3;6-4;0) also differed from the middle age group in the typical sample who were on average a year younger (2;6<3;0) ($z = -2.67$, $p < .008$).

Reflecting these group differences, a striking 48.8% of children in the clinic sample obtained scores more than 2 standard deviations below the typical mean for their age. On the other hand, a small proportion fared well relative to the typical group: 10.1% scored at or above the mean for their age, and 14.9% were within one standard deviation below the mean.

Item variables

Mean scores and standard deviations according to word status and item length are shown in Table 3. The ANOVA revealed a significant effect of word status, with an advantage for words over nonwords [$F(1,165) = 67.06$, $p < .001$, $\eta^2 = .29$]. In line with this, the majority of children repeated words more accurately than nonwords. Again, comparable results were obtained when the analysis was repeated for only those children who attempted every item [$F(1,98) = 36.48$, $p < .001$, $\eta^2 = .27$]. Taking the children who attempted at least 31 items on the test (82% of the sample), only

18.8% scored higher on nonwords, compared with 61.6% on words. Compared with the typical group, about 11% fewer children showed an advantage for words, a significant difference ($\chi^2 = 5.15$, df = 1, $p = .02$); about 5% more showed an advantage for nonwords, which was not significant ($\chi^2 = 1.98$, df = 1, $p = .16$). As with the typical group, the range of advantage for words (1-10) was notably greater than for nonwords (1-4).

The ANOVA also revealed a significant effect of item length [$F(2,330) = 233.42$, $p < .001$, $\eta^2 = .59$]. As was the case for the typical group, all post hoc comparisons for item length were significant ($p < .001$): repetition scores for one-syllable items were higher than repetition scores for two-syllable items, which in turn were higher than scores for three-syllable items. Contrary to the findings with the typical group, none of the two-way or three-way interactions were significant.

Syllable Loss and Prosodic Structure

Rate of syllable loss varied considerably, but was consistently higher than the rate of syllable loss in the typical sample. Just over a tenth (19 out of 168 children) never omitted or coalesced syllables (compared to a fifth in the typical sample). More than half the sample lost five or more syllables, compared with 22.5% of the typical sample. Seven children had higher loss of syllables than any child in the typical sample, with a maximum loss of 32.

Table 4 shows the mean and range of syllable loss according to prosodic position of the syllable, age, and item variables of lexical status and length. Given the floor effects, skewed distribution, and unequal variances of the data, a parametric analysis of these errors was not possible, so chi-square tests were used.

Table 4 about here

The mean rate of total syllable loss in the clinic sample, 11.9%, was roughly double that in the typical sample, 5.5% (χ^2 = 341.78, df = 1, p < .001). In contrast to this striking difference in rate of syllable loss, the profile of syllable loss was very similar across the two groups. As for the typical sample length had a highly significant impact on the level of loss ($loss_{2syllable}$ = 9.6%; $loss_{3syllable}$ = 14.5%; χ^2 = 42.19, df = 1, p < .001), as did age (χ^2 = 73.35, df = 2, p < .001). Likewise, prosodic structure had a significant impact on the rate of loss (χ^2 = 806.36, df = 2, p < .001). Again, unstressed syllables were more vulnerable than stressed syllables, and were by far the most vulnerable when they occurred before the stressed syllable, as illustrated in Figure 2.

Figure 2 about here

The groups only differed with respect to those syllable positions that were almost perfectly preserved in the typical group: primary stressed syllables in two- and three-syllable items, and poststress syllables in two-syllable items. Loss of syllables in these positions was still relatively low in the clinic group, but not negligible. So, while trochaic structures (strong syllable + weak syllable) were very robust in the typical sample, they were reduced by some children in the clinic sample, for example *ladder* → [jæd], *person* → [pɜt] /ˈdʒæmɪk/→ [dʒæm], [dʒʌk]. Syllables in relatively strong positions revealed a hierarchy of difficulty which was not observed in the typical sample: primary stressed syllables were better preserved in two-syllable than three-syllable items, which were in turn better preserved than post-stress syllables in two-syllable items. For the rest, the hierarchy was the same as that observed in the typical sample.

In contrast to the typical sample, lexical status did impact on the preservation of syllables in the clinic sample: words were significantly more vulnerable to loss overall than nonwords ($loss_{words}$ = 12.9%; $loss_{nonwords}$ = 10.9%; χ^2 = 7.81, df = 1, *p* = .005).

Discussion

Before discussing the results of the clinic study in relation to the study of typically developing children, it is important to recognise that these studies were not carried out in identical conditions. The typically developing children were assessed in their nursery, but conditions in different nurseries were not identical. The clinic children were assessed at home, but again, there were inevitable variations in conditions. The PSRep test was administered to the clinic group as part of a battery of assessments spread over two sessions. The typically developing children received only the PSRep and a receptive vocabulary test. Finally, as discussed above, presentation of the PSRep test to both groups was live and responsive to the child's attention and interaction, introducing variations in delivery of stimuli. Clearly, these conditions are not comparable to laboratory-based research. However, we need to consider the strengths of conducting research in children's natural environments. We would argue that the assessments were carried out in conditions that were comparable to clinical assessments and observations, were familiar to the child, and were most conducive to eliciting the maximum numbers and naturalness of responses. In addition, we would argue that the size of our samples and the consistency of patterns in the data give weight to our findings. Nevertheless, replication of studies using recorded presentation of stimuli is necessary before general conclusions can be drawn about levels and patterns of performance.

Starting with levels of performance, the scores of the clinic sample were found to be substantially lower than those of the typical sample; only the oldest clinic group, aged 3;6-4;0 attained scores of a similar order to the typical sample, and their profile was closest to the youngest typical group, on average 18 months younger. The extent of divergence is captured by the finding that nearly half of the clinically referred children obtained scores more than 2 standard deviations below the mean for their age. Nevertheless, a tenth of the clinic sample achieved scores at or above the mean for their age, and a quarter were within one standard deviation of the mean.

The differences observed between the typical and clinic samples in this study are broadly in line with the many studies that have found significant differences between children with SLI and typically developing children (Bishop, North, & Donlan, 1996; Botting & Conti-Ramsden, 2001; Conti-Ramsden, Botting, & Faragher, 2001; Dollaghan & Campbell, 1998; Edwards & Lahey, 1998; Ellis Weismer et al., 2000; Gathercole & Baddeley, 1990; Gray, 2003; Marton & Schwartz, 2003; Montgomery, 2004; Norbury, Bishop, & Briscoe, 2002). As in these studies, performance on the repetition test proved highly differentiating. Differences between the nature of the samples and the tests used in these studies preclude more detailed comparisons. First, our samples of children were younger than those in other studies. Second, and most importantly, our clinic sample was selected on the basis of early referral due to concerns about language, rather than identified language impairment, and would be expected to include some children with no language difficulties. Likewise, it is possible that the typical sample included children with unidentified language problems. Hence, we would expect some overlap between our typical and clinic samples.

The error measure we used, rate of syllable loss, also differentiated between clinic and typical samples, with the clinically referred children losing roughly twice the number of syllables lost by their peers. It is notable, though, that a minority of every clinic age band lost syllables at a rate very rarely observed in the typical sample.

In contrast to *levels* of performance, *patterns* of performance were strikingly similar across the clinic and typical groups. Both groups showed effects of prosodic structure of items on syllable loss, with vulnerability to syllable loss showing the same prosodic hierarchy. The most striking difference was that some children in the clinic sample lost syllables in prosodic positions that were virtually at floor in the typical group – syllables carrying stress, and syllables occurring poststress in two-syllable items. In such positions, the discrepancy between syllable loss in the two groups was as high as six-fold, even though the rate of syllable loss in these positions was still low relative to other prosodic positions. The finding of differences in absolute but not relative numbers of syllables lost in different prosodic positions suggests a delayed rather than atypical pattern of repetition in the clinic group.

In line with our predictions, both groups showed the same strong effects of item length, with reductions in performance as number of syllables increased. Both were also affected by lexical status, with children attempting more words than nonwords, and with a far greater number of children showing an advantage for words over nonwords than the reverse. In both groups the advantage for words (up to 11) greatly exceeded the advantage for nonwords (up to 4). It is striking that no child in the clinic sample showed a higher advantage for words or for nonwords than was observed in the typical sample. As pointed out above, effects of lexical status observed in our study may have been influenced by the neighborhood density of words and phonotactic probability of nonwords in the PSRep test, which may in turn have influenced the relation we observed between repetition performance and vocabulary. In future studies, it would be interesting to investigate whether typical and clinic groups show differential effects of these factors, and if so, whether the effects are correlated with vocabulary differences, as we would expect given correlations that have been found in typically developing children.

Such investigation may throw more light on two subtle differences between the typical and clinical samples in our study, both of which involved lexical status. First, rate of syllable loss was higher in words than nonwords for the clinic sample, but did not differ in the typical sample. Second, a smaller proportion of the clinic sample showed an advantage for words. Both findings point to a relatively greater vulnerability of real words in the clinic sample compared with the typical sample. This vulnerability could be due to children accessing weak representations, or representations in which earlier constraints on phonology are ‘frozen’; since nonwords are not stored, they would not be subject to any such limitations in children’s representations. These findings identify the need for future investigations into the role of representations and processing in these children’s repetition.

Apart from these subtle differences in effects of lexical status on the repetition performance of the clinic sample, we have seen that they presented a general picture of delay which was in line with delay in their language. However, a small number of children showed unusual profiles. Most notable was the occurrence of extreme mismatches between performance on the PSRep Test and on language assessments. Some children showed very poor repetition skills but attained language scores at or above the mean. Conversely, a few children attained repetition scores at or above the mean, but were low on assessments of language. These mismatches go against the tide of positive relations between repetition and language observed in this and many previous studies (see introduction). Turning to performance on the repetition test itself, some children showed an unusually high rate of syllable loss, in some cases affecting syllables that are prosodically very robust and almost invulnerable to loss in typically developing children. The longer-term outcome of these atypical profiles will establish whether they reflect particular underlying problems, and whether they are clinically informative.

One of the main motivations behind the development of the PSRep Test was the hypothesis that repetition skills would be predictive of later language abilities, and might serve as a clinical predictor. The findings of the current study, which subtantiate and extend our previously reported findings (Roy & Chiat, 2004), are prerequisites for investigation of its predictive potential. The study has demonstrated that the test is highly differentiating between our groups of typically developing and clinically referred children. It has also identified some atypical profiles within the clinic sample. The repetition performance of the sample as a whole was delayed, but some children attained scores within the normal and even high range, while a small number showed unusual rather than delayed patterns of response. The sample as a whole showed concurrent relationships with language as measured by a standardised language test and parental report, but a small number of children showed unusual mismatches between these. Discrimination between the clinic and typical samples and identification of variable profiles within the clinic sample are crucial if the PSRep Test is to have the potential to predict longer-term problems with language, and the nature of different children's problems.

The findings of this study lay the foundations for investigating whether poor levels of performance, unusual profiles of performance, and unusual mismatches between performance on repetition and language relate to later profiles. A follow-up study of the clinic cohort reported in this paper is currently underway. A range of assessments of language, pragmatics and phonological awareness have been carried out approximately 18 months later, when the children are aged 4-5 years. One aim is to determine whether early profiles on the PSRep Test alone, or in combination with other measures, both standardised and novel, are associated with later speech, language or pragmatic deficits. This follow-up study will reveal whether the PSRep Test may indeed serve as a useful clinical predictor.

Footnotes

[1] As Kolmagorav-Smirnov tests revealed that the distribution of scores for the sample as a whole and the two older groups (2;6<3;0; and 3;0-4;0) differed significantly from normality, equivalent nonparametric tests were conducted for all analyses and comparable results were achieved (full analyses available on request).

[2] Kolmagorav-Smirnov tests in the case of the clinic sample revealed that the distribution of scores for the sample as a whole and the youngest age group (2;6<3;0) differed significantly from normality. As was the case in the typical sample equivalent nonparametric tests were conducted for all analyses and comparable results were achieved (full analyses available on request).

Acknowledgements

The study of the clinic sample reported in this paper was funded by the Economic and Social Research Council, Award No. RES-000-23-0019. The study of the typical sample was funded by City University Pump Priming and Departmental Research Funds. Many thanks to Talia Barry, Alex Carter, Sophie Edgington, Melissa Humbert, Tracey Jennings, Renia Kaperoni, Luisa Martinez, Louise Occomore, Sophia Qureshi and Sharonne Williams for carrying out the assessments.

References

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.

Adams, A.-M., & Gathercole, S.E. (2000). Limitations in working memory: Implications for language development. *International Journal of Language and Communication Disorders, 35*, 95-116.

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, 37*, 391-403.

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

Boucher, J., & Lewis, V. (1997). *Pre-school Language Scale-3 (UK Adaptation)*. London: Harcourt Brace & Company Books.

Burt, L., Holm, A., & Dodd, B. (1999). Phonological awareness skills of 4-year-old British children: An assessment and developmental data. *International Journal of Language and Communication Disorders, 34*, 311-335.

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

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

Dale, P.S., Dionne, G., Eley, T.C., & Plomin, R. (2000). Lexical and grammatical development: A behavioural genetic perspective. *Journal of Child Language, 27*, 619-642.

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, Berkshire: NFER-Nelson.

Edwards, J., Beckman, M.E., & Munson, B. (2004). The interaction between vocabulary size and phonotactic probability effects on children's production accuracy and fluency in nonword repetition. *Journal of Speech, Language, and Hearing Research, 47*, 421-436.

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. (1996). *BAS II: British Ability Scales*. Second Edition. Windsor, Berkshire: NFER-Nelson.

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

Gathercole, S.E., & Adams, A-M. (1993). Phonological working memory in very young children. *Developmental Psychology, 29*, 770-778.

Gathercole, S.E., & Baddeley, A.D. (1989). Evaluation of the role of phonological STM in the development of vocabulary in children: A longitudinal 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.

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

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

Gathercole, S.E., Willis, C.S., Baddeley, A.D., & Emslie, H. (1994). The Children's Test of Nonword Repetition: A test of phonological working memory. *Memory, 2*, 103-27.

Gray, S. (2003). Diagnostic accuracy and test-retest reliability of nonword repetition and digit span tasks administered to preschool children with specific language impairment. *Journal of Communication Disorders, 36*, 129-151.

Hart, B., & Risley, T.R. (1995). *Meaningful differences in the everyday experience of young American children*. Baltimore: Brookes.

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

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

Munson, B., Edwards, J., & Beckman, M.E. (2005). Relationships between nonword repetition accuracy and other measures of linguistic development in children with phonological disorders. *Journal of Speech, Language, and Hearing Research, 48*, 61-78.

Munson, B., Swenson, C.L., & Manthei, S.C. (2005). Lexical and phonological organization in children: Evidence from repetition tasks. *Journal of Speech, Language, and Hearing Research, 48*, 108-124.

Norbury, C.F., Bishop, D.V.M., & Briscoe, J. (2002). Does impaired grammatical comprehension provide evidence for an innate grammar module? *Applied Psycholinguistics, 23*, 247-268.

Roy, P., & Chiat, S. (2004). A prosodically controlled word and nonword repetition task for 2- to 4-year-olds: Evidence from typically developing children. *Journal of Speech, Language, and Hearing Research, 47*, 223-34.

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

Appendix: List of Items

Practice items		**Words**	**Nonwords**
		nose /nəʊz/	/nɑz/
		button /ˈbʌtən/	/ˈtʌbən/

Test items

Length	**Stress**[1]	**Words**	**Nonwords**[2]
1-syllable		jar, toe, egg, arm, lamb, mouse	/dʒaɪ tɜ ʊg ɔm lɒm mis/
		/dʒɑ təʊ ɛg ɑm læm mɑʊs/	
2-syllable	ˈSW	ladder, person, magic	/ˈdælə ˈsɜpən ˈdʒæmɪk/
		/ˈlædə ˈpɜsən ˈmædʒɪk/	
2-syllable	WˈS	police, machine, balloon	/ləˈpis ʃəˈmin ləˈbun/
		/pəˈlis məˈʃin bəˈlun/	
3-syllable	ˈSWˌS	dinosaur, holiday	/ˈsaɪnəˌdɔ ˈlɒdɪˌheɪ/
		/ˈdaɪnəˌsɔ ˈhɒlɪˌdeɪ/	
3-syllable	WˈSW	banana, computer	/nəˈnɑbə tɒŋˈkjupə/
		/bəˈnɑnə kɒmˈpjutə/	
3-syllable	ˌSWˈS	magazine, cigarette	/ˌgæzəˈmin ˌrɪgəˈsɛt/
		/ˌmægəˈzin ˌsɪgəˈrɛt/	

1 Primary stress is indicated by ˈS; secondary stress is indicated by ˌS.

2 Transcription is for targets in Southern British Standard English.

Table 1. Typical sample: Word and nonword scores according to item length and age group

	Σ sample (n=315)		*2;0<2;6 (n=66)*		*2;6<3;0 (n=75)*		*3;0-4;0 (n=174)*	
Items	**M**	**SD**	**M**	**SD**	**M**	**SD**	**M**	**SD**
Word								
1 syllable	5.24	1.2	4.59	1.49	5.05	1.36	5.57	0.89
2 syllable	4.53	1.62	3.42	1.93	4.23	1.69	5.07	1.17
3 syllable	3.96	1.89	2.48	2.07	3.39	1.92	4.76	1.29
Total	13.86	4.19	10.55	4.83	12.88	4.23	15.55	2.83
Nonword								
1 syllable	4.84	1.50	4.17	1.80	4.63	1.79	5.20	1.11
2 syllable	3.84	1.84	2.86	1.86	3.52	1.86	4.36	1.63
3 syllable	2.97	1.74	1.98	1.71	2.56	1.77	3.53	1.51
Total	11.64	4.33	9.02	4.81	10.71	4.73	13.03	3.28
Word/Nonword								
Total	25.50	8.08	19.56	9.29	23.59	8 .48	28.58	5.55

Table 2. Typical sample: Percentage syllable loss according to prosodic structure, item length, age group, and lexical status

	Primary stress		***Secondary stress***	***Poststress***		***Prestress***	
Items	**2 syll.**	**3 syll.**	**3 syll.**	**2 syll.**	**3 syll.**	**2 syll.**	**3 syll.**
Age group							
2;0 < 2;6	0.30	2.13	5.47	1.47	15.05	30.56	45.21
2;6 < 3.0	0.24	0.86	3.86	0.72	10.04	20.88	26.74
3;0 – 4;0	0.05	0.34	1.40	0.39	4.70	12.43	14.33
Lexical status							
Words	0.17	0.79	2.20	0.89	8.95	15.21	23.15
Nonwords	0.11	0.80	3.27	0.45	6.77	20.50	22.76
Σ items	0.14	0.80	2.73	0.67	7.87	17.84	22.96

Table 3. Clinic sample: Word and nonword scores according to item length and age group

Items	***Σ sample (n=168)***		***2;6<3;0 (n=57)***		***3;0<3;6 (n=70)***		***3;6-4;0 (n=41)***	
	M	***SD***	***M***	***SD***	***M***	***SD***	***M***	***SD***
Word								
1 syllable	3.64	1.93	2.74	2.04	3.66	1.77	4.88	1.27
2 syllable	2.58	2.00	1.60	1.76	2.8	1.97	3.59	1.8
3 syllable	1.61	1.76	0.77	1.32	1.8	1.79	2.46	1.73
Total	7.84	5.05	5.11	4.58	8.26	4.79	10.93	4.13
Nonword								
1 syllable	3.17	1.91	2.21	1.83	3.27	1.74	4.34	1.58
2 syllable	1.92	1.79	1.14	1.66	1.97	1.67	2.93	1.66
3 syllable	1.16	1.43	0.53	1.17	1.26	1.45	1.88	1.38
Total	6.26	4.46	3.88	3.98	6.5	4.13	9.15	3.86
Word/nonword	14.1	9.22	8.98	8.34	14.76	8.53	20.07	7.55

Table 4. Clinic sample: Percentage syllable loss according to prosodic structure, item length, age group, and lexical status

	Primary stress		***Secondary stress***	***Poststress***		***Prestress***	
Items	**2 syll.**	**3 syll.**	**3 syll.**	**2 syll.**	**3 syll.**	**2 syll.**	**3 syll.**
Age group							
2;6 < 3;0	0.8	10.08	15.67	8.63	19.75	37.25	38.6
3;0 < 3.6	2.02	4.14	11.55	7.46	19.43	30.69	35
3;6 - 4;0	0.21	2.74	2.52	0.83	11.84	22.08	29.49
Lexical status							
Words	0.99	5.43	10.73	5.64	18.62	32.15	40.6
Nonwords	1.4	4.29	9.64	6.45	16.19	28.1	27.9
Σ items	1.18	4.88	10.2	6.03	17.43	30.18	34.49

Figure 1: Total repetition scores according to age for typical and clinic samples

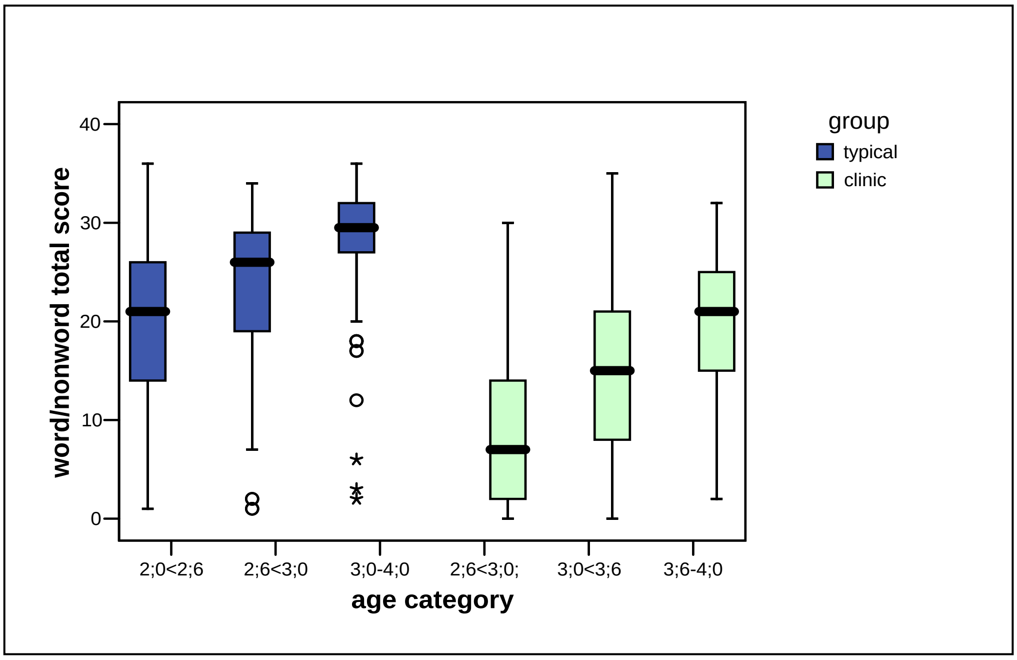

Figure 2: Percentage syllable loss according to prosodic position in typical and clinic samples