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(1), 223-234.

A prosodically-controlled word and nonword repetition task for 2-4 year olds: Evidence from typically developing children

Penny Roy

Department of Language and Communication Science

City University

Northampton Square

London EC1V 0HB

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

Shula Chiat

Department of Human Communication Science

University College London

Chandler House

2 Wakefield Street

London WC1N 1PF

e-mail: shula.chiat@ucl.ac.uk

Abstract

An association has been found between nonword repetition and language skills in school-aged children with both typical and atypical language development (Dollaghan & Campbell, 1998; Ellis Weismer et al., 2000; Gathercole & Baddeley, 1990; Montgomery, 2002). This raises the possibility that younger children's repetition performance may be predictive of later language deficits. In order to investigate this possibility, it is important to establish that elicited repetition with very young children is both feasible and informative. To this end, a repetition task was designed and carried out with 66 children aged 2-4. The task consisted of 18 words and 18 matched nonwords that were systematically manipulated for length and prosodic structure. In addition, an assessment of receptive vocabulary was administered.

The repetition task elicited high levels of response. Total scores as well as word and nonword scores were sensitive to age. Lexical status and item length affected performance regardless of age: words were repeated more accurately than nonwords, and one-syllable items were repeated more accurately than two-syllable items, which were in turn repeated more accurately than three-syllable items. The effect of prosodic structure was also significant. Whole syllable errors were almost exclusive to unstressed syllables, with those preceding stress being most vulnerable. Performance on the repetition task was significantly correlated with performance on the receptive vocabulary test. Since this repetition task was effective in eliciting responses from most of the 2 to 4-year-old participants, tapped developmental change in their repetition skills, and revealed patterns in their performance, it has the potential to identify deficits in very early repetition skills that may be indicative of wider language difficulties.

A prosodically-controlled word and nonword repetition task for 2-4 year olds: Evidence from typically developing children

The repetition of nonwords has gained remarkable status as a measure of children's language over the last decade. It started its research life as a tool developed by Gathercole and Baddeley (1989) to evaluate their hypothesis that phonological short-term memory (STM) plays an important role in vocabulary acquisition. Nonword repetition was assumed to provide a measure of phonological STM. This assumption has not gone unquestioned; it has been a catalyst for debate and research on what nonword repetition actually measures. However, the current high profile of nonword repetition testing is due more to the relationships that have been found between nonword repetition and a range of other language abilities and disabilities.

Nonword Repetition and Language Measures

A number of studies have revealed correlations between nonword repetition and receptive and expressive vocabulary size, as well as indices of speech output including repertoire of vocabulary, utterance length and grammatical complexity, in groups of typically developing children ranging from 3-5 years of age (Adams & Gathercole, 1995, 2000; Gathercole & Adams, 1993; Gathercole & Baddeley, 1989). Gathercole and Baddeley's (1990) investigation of nonword repetition in a small group of children with language disorders extended the evidence, revealing significant differences between the performance of these children and the performance of verbally and non-verbally matched controls. These findings have been corroborated by a growing number of studies. All have reported nonword repetition deficits and correlations between nonword repetition and measures of language in children with language impairment (e.g., Botting & Conti-Ramsden, 2001; Dollaghan & Campbell, 1998; Edwards & Lahey, 1998; Ellis Weismer et al., 2000; Montgomery, 1995, 2002; Norbury, Bishop & Briscoe, 2002; Stothard, Snowling, Bishop, Chipchase & Kaplan, 1998), and even in children whose language difficulties appear to have resolved (Bishop, North & Donlan, 1996; Conti-Ramsden, Botting & Faragher, 2001).

Nonword Repetition as a Potential Clinical Marker

Bishop et al. (1996) was the first study to establish that a deficit in nonword repetition is significantly heritable, leading to the suggestion that nonword repetition "provides a marker of the phenotype of developmental language impairment" (p. 156). Other indices of language have also been proposed as clinical markers, for example, children's production of finite verb forms. Bedore and Leonard (1998) found that a finite verb morpheme composite was highly accurate in discriminating children with SLI aged 3;7-5;9 (year;month) from age-matched controls, in keeping with the findings of Rice, Wexler and Cleave (1995). When Conti-Ramsden et al. (2001) compared four proposed clinical markers, they found that the overall accuracy of nonword repetition in identifying 11-year-olds with SLI was high – marginally poorer than sentence recall, but considerably better than past tense and third person singular marking.

It has been argued that a viable clinical marker should be largely independent of IQ (Bishop et al., 1996). Nonword repetition meets this criterion: a number of studies have found it to be largely independent of performance IQ in children with both typical and atypical language development (Conti-Ramsden et al., 2001; Ellis Weismer et al., 2000). A further advantage for nonword repetition is that it appears to be a culturally unbiased measure (Dollaghan & Campbell, 1998), unlike other measures of language that are known to be associated with socioeconomic factors (Hart & Risley, 1995). Burt, Holm and Dodd (1999) found no difference in nonword repetition performance between children from upper middle class and working class areas in the UK. Similarly, Ellis Weismer et al. (2000) found no difference between ethnic groups represented in their USA study, and point out “the potential usefulness of processing-based measures generally, and nonword repetition tasks specifically, in providing culturally nonbiased assessments of linguistic abilities” (p. 874). Less attention has been given to gender, but Burt et al.’s (1999) investigation found no gender differences. This contrasts reported gender effects on measures of early language development, particularly vocabulary (Fenson et al., 2000; Rescorla & Alley, 2001). In summary, nonword repetition has strong potential as a processing-based clinical marker, uncontaminated by factors known to affect performance on most knowledge-based measures of language.

Nonword Repetition as a Potential Predictor

The findings regarding the relationship between deficits in nonword repetition and language deficits have significant implications for development of children younger than those that have been studied. The bulk of evidence to date comes from typically developing children whose language development is virtually complete. The majority of studies with English-speaking children have used either the Children’s Test of Nonword Repetition (Gathercole & Baddeley, 1990), which is standardized for children aged 4-8, or a test designed by Dollaghan and Campbell (1998) that was originally carried out with language-impaired children aged 6;0 to 9;9 and their age-matched controls. There is very limited evidence from children under the age of 4. If nonword repetition is a marker for SLI in older children, we might reasonably expect it to be a predictor of language impairment in younger children. This could be particularly important in the assessment of language in the very early years, as it is currently difficult to diagnose language problems at an early age.

Follow-up studies of very young children who show delay or impairment in language development illustrate the difficulties of early diagnosis. Bishop and Edmundson (1987) followed up children who presented with impairment on a range of standardized language assessments, in order to investigate the predictiveness of these assessments. They found that performance on these tests at 4 years old was predictive of outcome at 5½, with a story re-tell task proving the most reliable discriminator. However, as they point out, their findings applied to a narrow range of ages: “one could not generalise [these findings] to 3-year-olds where inability to tell a story might have a different significance” (p.170).

Studies following up younger children identified as late talkers at 24-31 months have found some, but not all, early measures to be predictive of later language skills (Rescorla, Dahlsgaard, & Roberts, 2000; Rescorla, 2002). For example, Rescorla’s Language Development Survey at 2 to 3 years was significantly associated with language and reading scores at 6 to 9 years (Rescorla, 2002). However, while measures such as this afford good group predictive validity, they are less informative about individual outcomes and ‘clinical caseness.’ It may be that assessments tapping emergent skills key to language development will provide better individual and clinical prediction than assessments of language itself.

In the search for key emergent skills, repetition is a promising avenue to explore. It may provide a better indicator of which children with delayed language will have persistent and specific problems with language. In order to investigate the predictive value of repetition in younger children, we need a test that is applicable to younger children. The goal of the present study was to develop a repetition task for typically developing 2 to 4-year-old children, which shows differences within this age range. In designing such a test, it was important to take account of factors that may affect nonword repetition, and whose effects may be informative about children’s phonological processing.

Key Factors in Test Design

One factor known to be influential is item length. This has been systematically varied in existing tests of nonword repetition, and studies using these tests have consistently revealed that length affects performance (e.g., Dollaghan & Campbell, 1998; Gathercole, Willis, Baddeley, & Emslie, 1994).

Another factor known to affect nonword repetition is wordlikeness: nonwords that are more word-like are easier for children to repeat than nonwords that are less word-like (Gathercole et al., 1994; Gathercole & Martin, 1996). In the case of very young children whose exposure to words is relatively recent, it may be that children do not have sufficient familiarity with real words for these to influence performance on nonwords. In order to investigate effects of familiarity on repetition at this early stage, we compared actual words and phonologically matched nonwords, rather than comparing more and less word-like nonwords.

In contrast to length and wordlikeness, the prosodic structure of items has rarely been investigated. Yet it is well established that prosodic factors are particularly influential in early development. Echols (1996) highlights the effects of prosodic structure on syllable omission in early word production. Children acquiring English show a preference for trochaic foot structure, in which a stressed (strong) syllable is followed by an unstressed (weak) syllable (SW). They are far more likely to omit weak syllables when these precede a strong syllable (WS) and therefore fall outside this foot structure. For example, children will almost certainly preserve the unstressed syllable in the trochaic form *ˈtiger*, but they may well omit the unstressed syllable preceding stress in *guiˈtar*. Kehoe (1997) provides further evidence of children’s sensitivity to the prevalent stress patterns of English. In an investigation of 22 to 34-month-olds’ imitation of real and novel words, Kehoe found that stress errors were more frequent in items that had less typical stress patterns. Similar effects of prosodic factors have been observed in the imitation of function morphemes by normally developing children (Gerken, 1991, 1994) and children with SLI (McGregor & Leonard, 1994). Yet just one study of word and nonword repetition in children with language impairment has considered the effects of prosodic structure on syllable omission. In this study, Sahlϑn et al. (1999) examined the vulnerability of unstressed syllables in repetition of words and nonwords by Swedish 5-year-olds with language impairment. Omission of whole syllables was found to be rare overall, but interestingly, there were six times more omissions of pre stress syllables than post stress syllables.

In the light of these findings, prosody was considered an important factor in the design of a pre school repetition test. By systematically manipulating the prosodic structure of items in our task, we may expose patterns of strength and weakness in typically developing children, with some prosodic structures eliciting more, or more gross, errors than others. A common finding in children with SLI is that their language shows disproportionate effects of factors that influence normal development (Chiat, 2000; Johnston & Schery, 1976). If we extend this observation to nonword repetition, we might expect prosodically difficult structures to be disproportionately difficult for children with SLI.

Articulatory complexity is yet another factor that is bound to influence performance in the target age group, but one that we are not aiming to tap. In order to control for its effects, we have kept items as articulatorily simple as possible, for example, by avoiding clusters. However, the limitations of children's early vocabularies meant that we were not always able to achieve this. Finding items that met our prosodic requirements led us to include some that were more articulatorily complex than we would have liked, e.g. *computer* and its matched nonword *tonkyooper* /tɒŋˈkjupə/. In order to avoid effects of articulatory complexity in our analysis of children's performance, we have attempted to match the segmental content of items in different prosodic categories. More importantly, we have allowed for articulatory simplification in our scoring of children's repetition.

Taking these considerations into account, we developed a repetition test consisting of words and phonologically matched nonwords that are systematically varied for length and prosodic structure. This paper reports the performance of a group of 2 to 4-year-olds on this test, analysing the effects of one between-subject variable (age), and three within-subject variables (word status, length, prosodic structure). Our predictions were:

1. Age: 3-year-olds would perform significantly better than 2-year-olds.
2. Word status: Word scores would be significantly higher than nonword scores.
3. Length: Scores on one-syllable items would be significantly higher than scores on two-syllable items, which would be higher than scores on three-syllable items.
4. Prosodic structure: Unstressed syllables would be significantly more vulnerable to omission than stressed syllables, and unstressed syllables outside a trochaic foot (i.e., preceding stress) would be more vulnerable than those within a trochaic foot (i.e., following stress).

In addition, a test of receptive vocabulary was administered for comparison with performance on our word and nonword repetition test.

Method

Participants

The participants in the study were 66 children with English as a first language, no history of hearing loss, no speech and language or developmental difficulties, and no significant neurological or medical difficulties. All were recruited from nurseries or playgroups in central London and an urban area of southeast England. Some of these establishments were privately owned and funded by parents (those of 41 children), and some were state-run and subsidized (those of 25 children). Detailed information about socioeconomic status was not collected, but informal discussion with heads of nurseries confirmed that children in privately run nurseries were from professional families, while those in state-run nurseries were from lower socioeconomic classes.

Ten recruits were earlier lost from the study either because they were identified as having otitis media with effusion (2 children) or because they failed to complete the task (8 children). The final sample therefore reflected a non-cooperation rate of 11%.

The 66 participants were aged between 24 and 47 months (M = 35.9, SD = 5.2), divided into two age bands. There were 27 children in the younger band, ranging from 2;0 to 2;11, and 35 children in the older band, ranging from 3;0 to 3;11. Boys and girls were represented equally (n = 33 for each). Boys were slightly older than girls on average, but not significantly so (M_{boys} = 36.8, SD = 4.8; M_{girls} = 35.1, SD = 5.5; t(64) = 1.3, *p*=.2).

Ethical approval and parental consent for children's participation in the study were obtained prior to their inclusion.

Test items

The task consisted of 4 practice items and 36 test items, both containing equal numbers of words and nonwords (see Appendix for the full list). The test items were controlled for length and prosody.

Length. The 18 words and 18 nonwords included equal numbers of one-syllable, two-syllable and three-syllable items (6 of each).

Prosodic structure. Stress patterns of target words and nonwords were systematically manipulated as follows:

Two-syllable items: 3 with ˈSW structure, e.g., ˈ*magic*

3 with WˈS structure, e.g., *maˈchine*.

Three-syllable items: 2 with ˈSWS structure, e.g., ˈ*dinosaur*

2 with SWˈS structure, e.g., *magaˈzine*

2 with WˈSW structure, e.g., *baˈnana*.

Lexical status. Words and nonwords were phonologically matched. One-syllable nonwords were created by altering the vowel in the matched word whilst maintaining vowel length, e.g., *arm* ➜ *orm* /ɔm/, *lamb* ➜ *lomm* /lɒm/. Two- and three-syllable nonwords were created by reversing two consonants in the matched word, e.g., *dinosaur* ➜ *sinodaur* /ˈsaɪnədɔ/, *computer* ➜ *tonkyooper* /tɒŋˈkjupə/.

Scoring

Evaluation of our hypotheses was based on measures of repetition accuracy and measures of syllable loss.

Accuracy scores. Responses were scored as correct if they contained all phonemic segments of the target in the correct order, with no additional phonemes. Other studies have made allowance for articulatory difficulties, even where participants are older and likely to have more mature articulation than our participants (Edwards & Lahey, 1998; Gathercole et al., 1994). We made allowances for target phonemes that were likely to pose articulation difficulties for very young children (see Dollaghan & Campbell, 1998), and also for target phonemes that might be subject to sociolinguistic variation. Specifically, we accepted the following immature articulations:

a) gliding of /r/, allowing [ʋ] and [w] for target /r/ in *cigarette, rigasette* /rɪgəˈsɛt/

b) stopping of affricates, allowing [d] for /dʒ/ in *jar, magic, juy* /dʒaɪ/, *jamic* /ˈdʒæmɪk/

c) fronting of post-alveolar fricatives and affricates, allowing [s] for /ʃ/ in *machine, shameen* /ʃəˈmin/ and [z] for /dʒ/ in *jar, magic, juy* /dʒaɪ/, *jamic* /ˈdʒæmɪk/

d) other substitutions that the child made consistently, that is, in all relevant instances within the data (e.g., stopping /s/, /z/ or substituting these with [θ]).

We accepted the following possible sociolinguistic variants:

a) omission of /h/ in *holiday* and *lodihay* /ˈlɒdɪheɪ/, which is acceptable in certain London varieties

b) elision of /ə/ in *police, balloon*, allowing [plis] and [blun], which are acceptable realizations of these words in some varieties.

Finally, we accepted substitution of phonetic variants, e.g., dentals for alveolar targets.

Since our ultimate goal was to produce a clinically viable tool, we used a relatively simple score of whole items correct, together with a simple measure of error -- rate of syllable loss -- explained below. To check if scoring whole items lost information that would be captured by more fine-grained and stringent measures, we also scored items for number of phonemes that were correct subject to allowances for articulatory and sociolinguistic variations listed above (a ‘liberal’ phoneme score), and number of phonemes correct with no allowances made (a ‘conservative’ phoneme score).

Entering either of these phoneme scores into our statistical analyses made no difference to our results, and proved no more informative. Carlson, Miller, Wright and Thal (2002) made a similar comparison between more and less stringent phoneme measures and also found no significant differences between these. We have therefore opted for the simpler measure, scoring accuracy of whole items; all statistical analyses of accuracy are based on whole item scores.

Syllable loss. In addition to scoring accuracy, we scored errors that resulted in the loss of a syllable. Loss of a syllable was recorded where

a) a vowel was omitted with or without adjacent consonants, e.g., [pis] for *lepeese* /ləˈpis/, [ʃin] for *machine*, [ˈnɑnə] for *banana*, [ˈhɒldeɪ] for *holiday*

b) two syllables were coalesced, combining the consonant from one with the vowel from the other, e.g., [bun] for *balloon*, [ˈbɑnə] for *banana*, [min] for *machine*, [dʒæk] for *jamic* /ˈdʒæmɪk/.

However, elision of schwa in *police* and *balloon* was not counted as syllable loss even though the coalescence of the target consonants into a cluster resulted in loss of the unstressed syllable, on the grounds that this is an acceptable realization of these lexical targets (see above).

Procedure

All children were assessed individually in a quiet area of their nursery or play group. The assessment session lasted about 20 minutes. The session started with a standardized test of receptive vocabulary, the long form of the British Picture Vocabulary Scales (BPVS), (Dunn & Dunn, 1982). This was followed by the repetition task. Words and nonwords were presented in two separate blocks. Order of presentation of the word and nonword blocks was counterbalanced across participants, and items within each block were randomized.

Before beginning the task, each child was introduced to a puppet with a movable mouth and was asked to copy some words that the puppet was going to say. The nonwords were presented in the same way except, in this case, stimuli were described as ‘silly puppet words’ rather than words. Two practice trials were given prior to the presentation of both word and nonword items to familiarize the child with the task. 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.

Most studies of nonword repetition have presented audio-recorded stimuli in order to ensure consistency and eliminate any visual cues (Dollaghan & Campbell, 1998; Gathercole & Baddeley, 1990). However, use of recorded stimuli is problematic with the very young age group involved in our study. As in the Sahlϑn et al. study (1999), stimuli were presented live to elicit cooperation and maximize responses, and every effort was made to maintain consistency of presentation. The availability of visual cues was not an issue for us, since we were using repetition as a measure of phonological skills that might draw on visual as well as auditory information. Children's responses were audiotaped and subsequently transcribed. Two students trained and highly proficient in phonetic transcription transcribed and scored the recorded responses of just over a third of the sample to check interrater reliability.

Results

For all results, *p* values were two-tailed, and where appropriate, effect sizes (η^2) are reported.

Reliability

Interrater agreement. The level of agreement between first and blind raters' scores for the total set of 36 items and the sets of words and nonwords was high and of an acceptable level, with intraclass correlations of .91, .95, and .95, respectively.

Internal consistency. Internal consistency of the total set of 36 items was calculated for the whole sample (n=66). The coefficient α (equivalent for dichotomous data to the KR20 coefficient) was 0.86. This value falls within the acceptable range suggested by Streiner and Norman (1995) of an α above 0.70 (Nunally, 1978), but not higher than 0.9, which may indicate that some of the items are unnecessary. Removal of any single item in the scale made a negligible impact on the resultant α values (0.85 –0.86).

Non-responses. Non-responses to items were relatively infrequent, accounting for only 2% of all potential responses. Just over two-thirds of the sample gave responses to all items, i.e., did not produce a single non-response. For the remaining third, non-responses ranged in frequency from 1-10, with fewer than a tenth of the sample producing four or more. The likelihood of non-response occurrences was unrelated to type of nursery placement or gender. Perhaps unsurprisingly, their occurrence was somewhat more frequent in younger children, but even then failure to make any attempt at an item was rare. Of children under 3 years of age, 11.5% had four or more non-responses, compared to 5% of children aged 3 or over. The seven oldest children, aged 3;6-3;11, attempted all items.

Comparing non-responses across items, 26/36 targets produced at least 1 non-response, but only two targets produced the maximum number of four non-responses. One of these was the two-syllable word *ladder*, which was the first item in the test; the other was the three-syllable nonword *tonkyooper* /tɒŋˈkjupə/. Overall, nonwords were marginally more vulnerable than words, with 58% of non-responses occurring on nonwords and 42% on words.

Nursery Type and Gender

An ANOVA was conducted with type of nursery placement (2 levels: state, private) and gender (2 levels: boys, girls) as between-subject factors. Type of nursery placement, in terms of being either privately owned or state-run, was not significantly associated with children's performance on the repetition test, M_{state}=28.2, SD=4.2; $M_{private}$=26.8, SD=6.6; F(1,66)=.9, *p*=.4. Similarly the difference between repetition scores for boys and girls was negligible and nonsignificant, M_{girls}=26.8, SD=6.6; M_{boys}=27.9, SD=5.0; F(1,66)=.2, *p*=.7. Accordingly, both factors -- type of nursery placement and gender -- were excluded from subsequent analyses.

Overall Task Performance

The 36 items, 18 words and 18 nonwords, evoked a wide range of response levels in the sample of 2 to 4-year-old children as a whole (range 7-35). While no child correctly repeated all items, four children achieved a score of 35, i.e., close to ceiling. Ninety percent of the sample achieved total scores in the range 20–35. The range was wider for the 2;0-2;11 than the 3;0-3;11-year-old group ($range_{2;0-2;11}$: 7-34; $range_{3;0-3;11}$: 22-35).

A breakdown of scores by age, lexical status and item length is given in Table 1. A mixed design ANOVA was conducted with word status (2 levels: word, nonword) and item length (3 levels: 1, 2, 3 syllables) as within-subject factors, and age group (2 levels: 2;0-2;11, 3.0-3;11) as a between-subject factor.

Table 1 about here

Age

The older children, aged 3-4 years, scored significantly higher on the repetition task than the younger children, aged 2-3 years, $M_{2;0\text{-}2;11}$=24.4, SD=7.0; $M_{3;0\text{-}3;11}$=29.3, SD=3.8; F(1,64)=13.2, *p*=.001, η^2=.17. As noted above, scores for the younger group were more variable, yielding a standard deviation nearly twice that of the older group on all measures (see Table 1 and Figure 1).

Figure 1 about here

These findings are in line with our predictions, and with previous research demonstrating significant developmental advances in the repetition performance of older children (Gathercole & Baddeley, 1990).

Word Status and Item Length

The mean repetition scores and standard deviations for all items according to word status are shown in Table 1.

The ANOVA showed that word status of test items discriminated between children's repetition performance. As a group, children correctly repeated significantly more words than nonwords, F(1,64)=60.4, *p*<. 001, η^2 =.49, indicating a substantial effect of word status.

The majority of children repeated words more accurately than nonwords, with only a small proportion of the sample scoring higher on nonwords (9%). As can be seen in Table 1, children correctly repeated an average of three more words than nonwords, M=2.8, SD=2.8.

The bias towards more successful repetition of words was also evident in the range of scores, with a maximum difference score of 11 between words and nonwords, and a minimum of -3. Given these figures, it is not surprising that the distribution of word scores was negatively skewed, with just over a half of the sample repeating 15 or more of the total 18 words correctly, compared to just under a quarter of the children scoring in the same range on the nonword items.

A similar picture emerged from an item analysis of pairs of words and their nonword counterparts. In the majority of cases (15/18 pairs), word scores were higher than their nonword counterparts. For half the total number of item pairs (n=10) the difference between the repetition of words and the repetition of nonwords was significant using related t tests (*p* values in the range of $<.05$ to $<.001$); in contrast, there was no pair where the nonword score significantly exceeded the score for its word counterpart.

Mean repetition scores and standard deviations for all items according to item length are shown in Table 1. Item length was found to have a significant effect on children's repetition scores. An increase in the number of syllables was associated with a decline in the number of items correctly repeated, $F(2,128) = 53.9$, $p < .001$, $\eta^2 = .46$. Post hoc comparisons using Bonferroni corrections revealed that children's repetition of monosyllabic items was significantly better than their repetition of disyllabic items, which in turn was significantly better than their repetition of trisyllabic items ($p < .001$).

A significant interaction between item length and word status was also found, $F(2,128) = 22.5$, $p < .001$, $\eta^2 = .26$, as shown in Figure 2. Analyses of simple main effects showed that this was due to children's greater facility for repeating trisyllabic words than trisyllabic nonwords. On average the children repeated three-syllable words as accurately as two-syllable words, although both scores were significantly poorer than their repetition scores for monosyllabic words. This comparability in repetition performance between word scores for two- and three-syllable items was not found for nonwords. The children found three-syllable nonwords significantly more difficult to repeat than two-syllable nonwords.

Figure 2 about here

No significant interaction was found between age and word status, $F(1,64)=.46$, $p=.5$, or between age and item length, $F(2,128)=.76$, $p=.4$. Similarly the three-way interaction between syllable length, word status and age was nonsignificant, $F(2,128)=.1$, $p=.8$.

Syllable Loss and Prosodic Structure

Loss of whole syllables was relatively rare. Twenty-five of the 66 children never omitted or coalesced syllables. Of the remaining 41 children, the majority lost no more than four syllables across their entire set of responses. Only 10 children omitted five or more syllables, and the maximum number of syllable omissions was 14 (two children).

Loss of syllables was analysed in relation 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 2 shows the total number of syllables lost as a percentage of the number of syllables in items that children attempted -- excluding non-responses -- according to their prosodic position. It also gives a breakdown of syllable loss by age, lexical status, and length.

Table 2 about here

Analysis of total percentage syllable loss reveals dramatic effects of prosodic structure, $\chi^2(3)=262.9$, $p<.001$. The overwhelming majority of losses involved unstressed syllables. Stressed syllables were preserved almost without exception, and secondary stressed syllables were relatively robust. Unstressed syllables were not. Those which occurred post-stress, following a strong syllable in a trochaic structure, were more liable to omission than stressed syllables (see Table 2). Unstressed syllables which occurred pre-stress, outside a trochaic structure, were by far the most vulnerable; they were about three times more liable to omission than post stress syllables, and 40 times more liable to omission than stressed syllables.

The rate of syllable loss changed with age (see Table 2). The younger group lost just over twice as many syllables as the older group, χ^2 (1) =34.4, p<.001. However, the relative vulnerability of different prosodic positions did not change with age. What changed is that vulnerable positions became less vulnerable with age. As shown in Table 2, the percentage loss of unstressed syllables, both pre- and post-stress, is two to three times higher in the younger age group.

As is evident from Table 2, rate of syllable loss was not affected by lexical status, χ^2 (1) = .77, p=.4, but was affected by item length, χ^2 (1) =17.6, p<.001, with significantly more loss in three-syllable than two-syllable items. However, the effect of length on syllable loss depended on prosodic structure. For example, post stress syllables were more vulnerable in three-syllable items than two-syllable items (7.2% versus 0.8% loss): in two-syllable items they were as robust as stressed syllables. However, post stress syllable were still not as vulnerable as pre stress syllables in two-syllable items (7.2% versus 12.7%). This rate of loss for pre stress syllables almost doubled in three-syllable items, rising from 12.7% to 22.1%.

Association with Receptive Vocabulary

Raw scores rather than standard scores from the test of receptive vocabulary (BPVS) were used in analysis, 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.

Repetition scores for all items, and for words and nonwords separately, were significantly associated with receptive vocabulary (BPVS) raw scores and age (total items: r_{voc}=.55, p<.001, r_{age}=.55, p<.001; words: r_{voc}=.53, p<.001, r_{age}=.58, p<.001; nonwords: r_{voc}=.45, p<.001, r_{age}=.42, p<.001). The correlation between age and receptive vocabulary was of a similar order (r=.53, p<.001). Correlations between repetition and receptive vocabulary scores with age partialled out yielded correlations of r=.36, r=.33 and r=.29 for total items, words and nonwords, respectively. These were statistically significant, with *p* values between .02 and <.001, so that repetition performance accounted for about 10% of the variance in receptive vocabulary performance. Stepwise regression analyses, with age entered first, showed that age and overall repetition scores accounted for 36% (adjusted R^2) of the variance in the receptive vocabulary scores (p<.001). Word and nonword scores, together with age, separately accounted for a very similar amount of variance in vocabulary scores: 34% and 33% respectively (p<.001). Overall repetition scores accounted for marginally more change in receptive vocabulary scores than age (β_{items} =.36, t=3.03, p=.004; β_{age} =.33, t=2.79, p=.007), whereas age accounted for either a comparable amount of change in vocabulary scores or slightly more than repetition scores for word and nonword separately (β_{word} =.34, t=2.74, p=.008; β_{age} =.34, t=2.75, p=.008; $\beta_{nonword}$ =.28, t=2.45, p=.02; β_{age} =.42, t=3.71, p<.001).

Discussion

The repetition task achieved a relatively good response rate, with loss of only 11% of children due to non-cooperation. For those children who cooperated, non-responses to individual items were also relatively rare. Even in the most extreme case, responses were obtained for nearly three-quarters of the test items. Furthermore, the level of interrater reliability and internal consistency was high. This confirms that repetition is a realistic and reliable method of assessment with children at this very early age.

Our initial analysis showed that nursery status and gender did not affect performance. Given the relatively small sample size, we are cautious about drawing conclusions about these factors from our study alone. However, our findings are in line with previous studies, which have found that nonword repetition is relatively independent of demographic factors (Burt et al., 1999; Dollaghan & Campbell, 1998; Ellis Weismer et al., 2000), an advantage for any clinical assessment tool.

It was crucial to the viability of the task to establish that repetition scores were sensitive to age. Our results show that this was the case. As predicted, average performance was significantly higher in the 3 to 4-year-old group than in the 2 to 3-year-old group. This confirms that our repetition task taps developmental change and has potential as an assessment tool for identifying children whose repetition skills are inappropriate for their age.

Interestingly, the performance of children in the younger age group was not only poorer on average, but showed considerably more variability than performance in the older age group. This matches Bates, Dale and Thal’s (1995) finding that variability in vocabulary development was particularly marked up to 30 months of age. They also noted the stability of individual variation over this period. It may be concluded that our task is particularly sensitive to skills that are emerging between 2 and 3 years, a key period in language development. This is essential for a task that was developed as a potential predictive tool for late talkers, who are typically identified during this age range.

The three linguistic variables also turned out to be significant, as predicted. The effect of lexical status was particularly striking, given the age group. Children whose exposure to words is relatively short-lived may not have benefited from the familiarity of lexical items. Contrary to this expectation, we found a significant difference between words and nonwords in both age groups, indicating that words already have an advantage. This is confirmed by the skewed distribution of differences between words and nonwords across the sample of children. Overall, 91% showed an advantage for words, with a mean difference score of 2.8 for the whole group. Furthermore, differences in favor of words extended beyond differences in favor of nonwords: individual children's word scores exceeded their nonword score by up to 11, while no child's nonword score exceeded his/her word score by more than three.

The effect of item length is less surprising, and consistent with findings of length effects in older children (Gathercole et al., 1994). With this younger age group, differences occurred between all item lengths. Interestingly, length effects were stronger for nonwords than words. This indicates that length effects interact with other factors, in this case, familiarity of the item. Similarly, Gathercole et al. (1994) found that length effects could be overridden, as evidenced by their finding that children performed slightly but consistently better on five-syllable nonwords than four-syllable nonwords. They attributed this to the greater morphological familiarity of the five-syllable items.

Length effects can also be influenced by prosodic factors, as our analysis of syllable loss demonstrated. Items of the same length were not equally vulnerable to syllable loss: rate of syllable loss depended on the prosodic structure of the item. In two-syllable items, an unstressed syllable that occurred pre stress was roughly 16 times more liable to loss than one occurring post stress. A similar, though smaller, discrepancy was found in three-syllable items, with loss of pre stress syllables roughly three times that of post stress syllables. Even more striking is the finding that prosodic factors sometimes overrode length effects. This is illustrated by our finding that loss of a pre stress syllable in two-syllable items was nearly double that of a post stress syllable in three-syllable items. These prosodic effects are entirely consistent with the widely observed effects of stress on early word production (e.g., Echols, 1996), and with the specific effects of stress on both word and nonword repetition reported by Sahlϑn et al. (1999). As in the Sahlϑn et al. study, we found that prosodic factors had similar effects on words and nonwords. The prosodic effects we have observed confirm the importance of taking prosody into account in selecting stimuli for tasks evaluating children's repetition skills.

Finally, scores on our repetition task were found to be significantly correlated with scores on the BPVS, a test of receptive vocabulary. A regression analysis showed that both word and nonword repetition accounted for a significant amount of change in receptive vocabulary. This mirrors Gathercole et al.'s (1994) finding that scores on nonword repetition were associated with concurrent receptive vocabulary scores. This association was arguably a driving force behind subsequent research on concurrent and predictive relationships between nonword repetition and other language abilities. Replicating the finding with 2 to 4-year-olds suggests that repetition skills are associated with vocabulary even at this young age, and supports the concurrent validity of our task. It should be noted, however, that the evidence is not unequivocal. For example, Edwards and Lahey (1998) found that nonword repetition was correlated with expressive vocabulary, but not with receptive vocabulary, in a group of children with SLI. On the other hand, Carlson et al. (2002) report the reverse for a group of typically developing children: correlations between nonword repetition and receptive vocabulary were of a similar order to ours, while correlations with expressive vocabulary were lower and nonsignificant. These relationships clearly need further investigation, particularly with respect to expressive vocabulary. A longitudinal study would clarify the direction of any relationships observed between nonword repetition and vocabulary.

Conclusions and Further Implications

The above summary of our findings shows that children as young as 2;0 are amenable to repetition testing. It also shows that measures of whole item score and whole syllable loss reveal differences between participants and effects of linguistic variables, without recourse to more time-consuming counts (e.g., number of phonemes correct). These characteristics make our test clinically realistic.

The effects we have observed indicate that our task is informative about language-related processing skills. Our findings support previous evidence that repetition is relatively independent of social class and gender, making it a potentially valuable test tapping processing skills rather than acquired knowledge. Clinically, the task provides a tool for identifying children whose repetition performance is poor overall compared to that of normally developing children. The systematic manipulation of lexical status, length, and prosodic structure of items in the task amplifies the information it can provide. It may reveal whether poor performance mirrors typical patterns, exaggerates these patterns, or is characterised by atypical patterns. Different patterns of performance could be of diagnostic significance.

On the basis of this relatively small-scale study, we can conclude that our task is a viable pre school assessment of repetition skills. It therefore provides the tool we need to address our longer-term goal of evaluating the clinical potential of word and nonword repetition. We are now using this task to investigate whether early repetition skills are predictive of later language development, and whether the task may serve as a very early clinical marker of language impairment.

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.

Bates, E., Dale, P.S., & Thal, D. (1995). Individual differences and their implications for theories of language development. In P. Fletcher & B. MacWhinney (Eds.), *The Handbook of Child Language* (pp. 96-151). Oxford, UK: Blackwell.

Bedore, L.M., & Leonard, L.B. (1998). Specific Language Impairment and grammatical morphology: A discriminant function analysis. *Journal of Speech, Language, and Hearing Research*, *41*, 1185-1192.

Bishop, D.V.M., & Edmundson, A. (1987). Language-impaired 4-year-olds: Distinguishing transient from persistent impairment. *Journal of Speech and Hearing Disorders, 52*, 156-173.

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.

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.

Carlson, J., Miller, S., Wright, C., & Thal, D. (2002, July). *Nonword repetition and language development in four-year-old children with and without a history of early language delay*. Poster session presented at the Joint Conference of the IX International Congress for the Study of Child Language and the Symposium on Research in Child Language Disorders, Madison, WI.

Chiat, S. (2000). *Understanding Children with Language Problems*. Cambridge, UK: Cambridge University Press.

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

Dunn, L., & Dunn, L. (1982). *The British Picture Vocabulary Scale*. Windsor, UK: NFER-Nelson.

Echols, C.H. (1996). A role for stress in early speech segmentation. In J.L. Morgan & K. Demuth (Eds.), *Signal to syntax: Bootstrapping from speech to grammar in early acquisition* (pp. 151-170). Mahwah, NJ: Lawrence Erlbaum Associates.

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.

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.

Fenson, L., Pethick, S., Renda, C., Cox, J.L., Dale, P.S., & Reznick, J.S. (2000). Short-form versions of the MacArthur Communicative Development Inventories. *Applied Psycholinguistics, 21*, 95-115.

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

Gathercole, S.E., and 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., & Martin, A.J. (1996). Interactive processes in phonological memory. In S.E. Gathercole (Ed.), *Models of Short-term Memory* (pp. 73-100). Hove, East Sussex, UK: Psychology Press.

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.

Gerken, L.A. (1991). The metrical basis for children's subjectless sentences. *Journal of Memory and Language, 30*, 431-451.

Gerken, L.A. (1994). Young children's representation of prosodic phonology: Evidence from English-speakers' weak syllable productions. *Journal of Memory and Language, 33,* 19-38.

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

Johnston, J.R., & Schery, T.K. (1976). The use of grammatical morphemes by children with communication disorders. In D.M. Morehead & A.E. Morehead (Eds.), *Normal and Deficient Child Language* (pp. 239-259). Baltimore: University Park Press.

Kehoe, M. (1997). Stress error patterns in English-speaking children's word productions. *Clinical Linguistics & Phonetics, 11*, 389-409.

McGregor, K.K., & Leonard, L.B. (1994). Subject pronoun and article omissions in the speech of children with specific language impairment. *Journal of Speech and Hearing Research, 37*, 171-181.

Montgomery, J. (1995). Sentence comprehension in children with specific language impairment: The role of phonological working memory. *Journal of Speech and Hearing Research, 38*, 187-199.

Montgomery, J.W. (2002). Understanding the language difficulties of children with specific language impairments: Does verbal working memory matter? *American Journal of Speech-Language Pathology, 11*, 77-91.

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.

Nunnally, J.C., Jr. (1978). *Psychometric theory* (2nd ed.). New York: McGraw-Hill.

Rescorla, L. (2002). Language and reading outcomes to age 9 in late-talking toddlers. *Journal of Speech, Language and Hearing Research, 45*, 360-371.

Rescorla, L. & Alley, A. (2001). Validation of the Language Development Survey (LDS): A parent report tool for identifying language delay in toddlers. *Journal of Speech, Language and Hearing Research, 44*, 434-445.

Rescorla, L., Dahlsgaard, K., & Roberts, J. (2000). Late-talking toddlers: MLU and IPSyn outcomes at 3;0 and 4;0. *Journal of Child Language, 27*, 643-64.

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

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

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*, 407-418.

Streiner, D.L., & Norman, G.R. (1995). *Health measurement scales: A practical guide to their development and use* (2nd ed.). Oxford: Oxford University Press.

Tables and Figures

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

items		*Σ sample (n=66)* *M*	*SD*	*2;0-2;11 (n=27)* *M*	*SD*	*3;0-3;11 (n=39)* *M*	*SD*
word:	1 syllable	5.4	.8	5.0	1.0	5.7	.5
	2 syllable	4.8	1.5	4.2	1.9	5.2	.9
	3 syllable	4.8	1.5	4.2	1.9	5.2	.9
	total	15.1	3.2	13.5	4.0	16.2	2.0
nonword:	1 syllable	5.1	1.0	4.7	1.1	5.3	.9
	2 syllable	4.2	1.4	3.7	1.7	4.5	1.1
	3 syllable	3.0	1.7	2.5	1.8	3.4	1.5
	total	12.3	3.3	11.0	3.9	13.5	2.4

2 tailed, p(age)=.001
2 tailed, p(word status) <.001
2 tailed, p(length) <.001
2 tailed, p(status.length)<.001

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

	Primary stress		*Secondary*	*Post-stress*		*Pre-stress*	
Items	2 syll	3 syll	3 syll	2 syll	3 syll	2 syll	3 syll
Age group							
2;0 - 2;11	0	1.6	3.8	1.3	11.8	18.7	30.8
3;0 - 3;11	0	0.2	2.0	0.4	4.1	8.7	16.2
Σ sample	0	0.8	2.7	0.8	7.2	12.7	22.1
Lexical status							
words	0	0.5	2.3	1.1	7.0	10.2	23
nonwords	0	1	3.1	0.5	7.5	15.3	21.1
Σ items	0	0.8	2.7	0.8	7.2	12.7	22.1

2 tailed, p(stress)<.001
2 tailed, p(length)<.001
2 tailed, p(age) <.001
2 tailed, p(status) =.4 n.s.

Figure 1. Boxplots showing total word and nonword scores according to age group

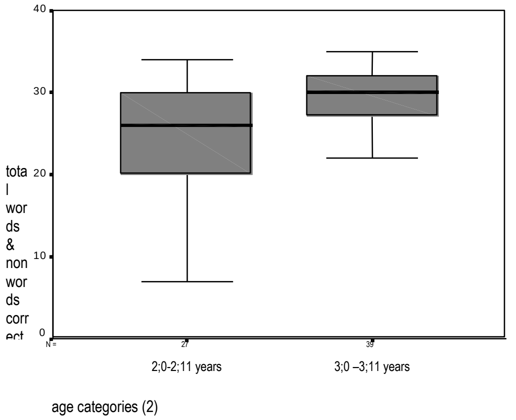
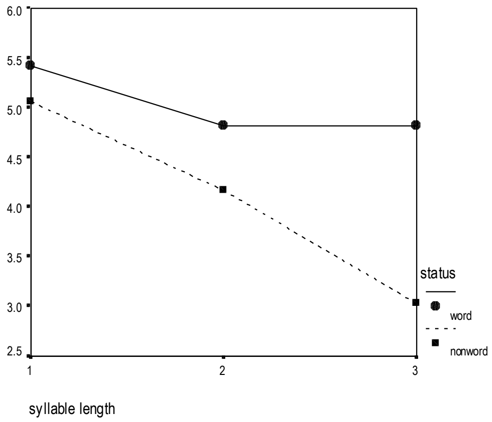


Figure 2. Word and nonword scores for 1, 2 and 3 syllable length items



Appendix: List of Test Items

Length, stress:	**Words**	**Nonwords**[1]
1-syllable:	jar, toe, egg, arm, lamb, mouse	juy, tur, ug, orm, lomm, meese
		/dʒaɪ, tɜ, ʊg, ɔm, lɒm, mis/
2-syllable, ˈSW:	ladder, person, magic	daller, serpen, jamic
		/ˈdælə, ˈsɜpən, ˈdʒæmɪk/
2-syllable, WˈS:	police, machine, balloon	lepeese, shameen, leboon
		/ləˈpis, ʃəˈmin, ləˈbun/
3-syllable, ˈSWS:	dinosaur, holiday	sinodaur, lodihay
		/ˈsaɪnədɔ, ˈlɒdɪheɪ/
3-syllable, WˈSW:	banana, computer	nanaaba, tonkyooper
		/nəˈnɑbə, tɒŋˈkjupə/
3-syllable, SWˈS:	magazine, cigarette	gazameen, rigasette
		/gæzəˈmin, rɪgəˈsɛt/

[1]Transcription is for targets in Southern British Standard English.