Nonword Repetition and Language Development in 4-Year-Old Children With and Without a History of Early Language Delay

Donna J. Thal
San Diego State University,
San Diego, California

Scott Miller
University of Wisconsin—Madison

Janna Carlson
Accent on Speech, San Diego, California

Martha Moreno Vega
San Diego State University,
San Diego, California

Purpose: This study examined the usefulness of the Nonword Repetition Test (NRT; C. Dollaghan & T. F. Campbell, 1998) with 4-year-old children and the relationship among the NRT, language, and other aspects of mental processing.
Method: The NRT was administered to 64 children at 4 years of age; 44 had a history of typical language development (HTD), and 20 had a history of language delay (HLD) at 16 months of age. Study 1 compared methods of scoring phoneme errors to determine whether the NRT was appropriate for this age group. Study 2 examined whether the NRT differentiated HTD from HLD. Study 3 examined the relations among scores on the NRT and standardized tests of language and mental processing.
Results: The NRT was found to be appropriate for 4-year-old children. Although all children had normal language abilities at the time of the study, the NRT (and several aspects of language and mental processing) differentiated between HTD and HLD. Relations among the NRT and other measures of language and mental processing were different from those previously reported, an unexpected finding that is inconsistent with traditional accounts of working memory and its relation to language development. Potential explanations are explored, and some directions for future research are suggested.

KEY WORDS: nonword repetition, language development, language delay, working memory

Processing models of language development and language disorders are being used more widely in the attempt to understand the nature of language ability. The working memory model originally proposed by Baddeley and colleagues (Baddeley & Hitch, 1974) has probably been the most broadly applied processing model in studies of adults (e.g., Baddeley, 2003; Baddeley, Chincotta, Stafford, & Turk, 2002; Cocchini, Logie, Della Sala, MacPherson, & Baddeley, 2002), school-age children (e.g., Alloway, Gathercole, Willis, & Adams, 2004; Dollaghan & Campbell, 1998; Gathercole & Pickering, 2000; Simkin & Conti-Ramsden, 2001), preschoolers (e.g., Avons, Wragg, Cupples, & Lovegrove, 1998; Bowey, 2001; Gathercole, Service, Hitch, Adams, & Martin, 1999), and children with language disorders (e.g., Botting & Conti-Ramsden, 2001; Briscoe, Bishop, & Norbury, 2001; Ellis Weismer et al., 2000; Jarrold, Baddeley, & Phillips, 2002). In that model, working memory comprises three subcomponents: (a) the central executive, (b) the visuospatial sketchpad, and (c) the phonological loop. The central executive controls the last two subcomponents and allocates attentional resources. The visuospatial sketchpad operates on visuospatial images, whereas the phonological loop functions to store phonological information and produce inner speech through an articulatory control center (Baddeley, 1992). Recently, a fourth component, called the *episodic buffer,* has been added to the central executive component of the model to help explain results in which visuospatial and phonological information have been found to be related in working memory (Baddeley, 2000, 2003). Baddeley and colleagues (Adams & Gathercole, 2000; Baddeley, Logie, & Ellis, 1988; Gathercole & Baddeley, 1990b) have proposed that differences in the function of the phonological loop may result in variations in the efficiency and accuracy with which stable phonological representations are stored in long-term memory and that these variations may provide an explanation of the individual differences seen in language learning, including the language learning problems found in children with specific language impairment (SLI).

Most of the developmental work based on this model has been carried out in the United Kingdom using a nonword repetition measure developed specifically for use with children: the Children's Nonword Repetition Test (CNRep; Gathercole, Willis, Baddeley, & Emslie, 1994). The test consists of 40 nonwords (e.g., *blonterstaping, fennerizer*) with 10 each at lengths of two, three, four, and five syllables, and the score received is the number of words repeated accurately. Research with this test has consistently shown links between nonword repetition scores and the development of language (Adams & Gathercole, 1995, 2000; Avons et al., 1998; Gathercole & Adams, 1993, 1994; Gathercole & Baddeley, 1990a, 1990b; Gathercole, Willis, et al., 1994). Children with SLI have been reported to perform particularly poorly on the CNRep (Bishop, North, & Donlan, 1996; Botting & Conti-Ramsden, 2001; Gathercole & Baddeley, 1990a) and similar nonword tests (Edwards & Lahey, 1998; Kamhi & Catts, 1986; Kamhi et al., 1988; Montgomery, 1995a, 1995b). However, interpretation of results varies considerably, especially with regard to the functional separation of working memory and language knowledge and whether differences are due to variations in the functioning of the phonological loop, an auditory perceptual processor, or a more generalized cross-domain processing mechanism. In addition, reports indicate that some children with resolved language disorders demonstrate significant impairment on the CNRep (Bishop et al., 1996), whereas some children with SLI do not (Botting & Conti-Ramsden, 2001). Thus, despite the connections between language skills and nonword repetition task scores reported in studies of typically developing children, it is possible to achieve normal language skills with poor nonword repetition abilities and to develop poor language skills when nonword repetition abilities are good. These findings suggest the need to consider complex interactional factors in the search for explanations of language development. We attempt to approach this issue in the research described in this article.

The design of the CNRep presents some problems for resolving the issue of separation between working memory and language knowledge, because it is not a pure test of working memory. It has been shown, for example, that the more wordlike nonwords are, the more they tap existing language knowledge (Dollaghan, Biber, & Campbell, 1993, 1995; Gathercole, 1995). The CNRep is not free of the word-likeness confound: All but one of the nonwords contain at least one root morpheme that corresponds to a real word, and some of the nonwords have bound morphological endings (e.g., *blonterstaping, contramponist*). Additional confounds with language knowledge and articulatory difficulty are created by the use of phonemes that are typically learned later (e.g., *thickery, loddernapish*) and consonant clusters (e.g., *trumpetine, stopograttic*).

Dollaghan and Campbell (1998) designed a test for speakers of American English that, because of more stringent constraints on the stimuli, minimizes the influence of language knowledge on nonword repetition. The Nonword Repetition Test (NRT; Dollaghan & Campbell, 1998) consists of 16 nonwords, 4 each at lengths of one, two, three, and four syllables (e.g., *naib, tevak, doitauvab, vetachaidoip*). All of the nonwords begin and end with consonants, none contain consonant clusters, none of the individual syllables correspond to English words, none of the words contain the "Late Eight" consonants (Shriberg & Kwiatkowski, 1994), and none of the consonants or vowels appears more than once in a particular nonword. The nonwords also contain only tense vowels (and therefore no weak syllables), and consonants occupy positions of low occurrence in real words (25%, following percentage-of-occurrence data from Shriberg and Kent (1982). Together, these constraints limit the effects of articulatory difficulty, perceptual difficulty, and language knowledge related to vocabulary, grammar, predictability of consonants, and stress patterns. The score used for the NRT is the total percentage of phonemes correctly imitated (TPPC) across all phoneme levels.

The NRT has been shown to differentiate between school-age children with and without language impairment (Dollaghan & Campbell, 1998; Ellis Weismer et al., 2000). However, it has not, to our knowledge, been used in studies designed to explore whether nonword repetition ability relies on an encapsulated, domain-specific module (e.g., a module dedicated to processing only phonological information) or a more general processor (e.g., one that processes all auditory information, or one that processes all temporally ordered input, regardless of perceptual domain). In addition, the usefulness of the test with children under age 6 years has not been examined.

In this article, we report the results of three studies. In Study 1, Dollaghan and Campbell's (1998) NRT was administered to 4-year-old children with language skills in the normal range to determine whether it could be used successfully with that age group. We compared three different scoring systems: the one used by Dollaghan and Campbell and two others in which phonemes that typically developing 4-year-old children may not normally produce were removed. Our hypothesis was that 4-year-old children with normal language would be able to complete the NRT and that the original scoring system would be as appropriate as either of the alternatives. In Study 2, we compared 4-year-old children with a history of typical language development to children who were the same age but who had a history of early language delay. On the basis of reports that older children with a history of language impairment who no longer scored in the impaired range in language did more poorly than children with a history of typical development on the CNRep (Bishop et al., 1996, Snowling, Chiat, & Hulme, 1991), we hypothesized that the children with a history of early language delay would score significantly more poorly than children with a history of typical language development on the NRT. In Study 3, we explored the relations between scores on the NRT and scores on standardized tests of language and mental processing for the combined group of typically developing children (regardless of early language development history, as in Study 1). We hypothesized that NRT scores would be closely related to vocabulary, grammar, and sequential mental processing abilities. This was partially a test of the Baddeley model claims. If nonword repetition tasks specifically tap the phonological loop, and the phonological loop is necessary for language learning (particularly for the development of new abilities within the auditory modality), then NRT scores (TPPC) should enter into a factor that includes vocabulary knowledge throughout language development and higher level language abilities as they are developed (Adams & Gathercole, 1995, 2000; Baddeley & Hitch, 2000). It was also an exploration of emergentist models (e.g., Bates & Goodman, 1997; Elman et al., 1996) in that nonword repetition, vocabulary, and grammar, like sequential processing abilities, may all be described as analytical skills that are mediated by the left hemisphere and that have changing interactions with the environment over developmental time.

Study 1

Method

In the first study, we compared three different methods of scoring the NRT (Dollaghan & Campbell, 1998) to determine whether typically developing children as young as 4 years of age could successfully complete the test and whether eliminating phonemes that may not be produced by all typically developing 4-year-old children would affect scores. Results were expected to provide evidence for the usefulness of the NRT at this younger age.

Participants. The participants were sixty-four 4-year-old children with language skills in the normal range who were part of a larger, longitudinal study of language development. The cohort of 37 males and 27 females had a mean age of 4;2 (years;months: range 4;0–4;6). All participants were originally recruited when they were 10 months old in the city of San Diego, California, and surrounding suburbs through local pediatric practices, parent bulletin boards on the Internet, flyers distributed at local day care centers, and friends of participants. All of the children were born at term, had a 5-min Apgar score of at least 7, and were from English-speaking families. The children had participated in laboratory evaluations four times previously, when they were 17, 21, 29, and 36 months of age. At each of those laboratory visits, the children received a pure tone hearing screening at 25 dB HL (American National Standards Institute) and an oral motor evaluation. At each of these four data points the oral motor evaluation involved a subjective evaluation of the lips, tongue, and jaw while the children were eating and drinking. All children were judged to have normal use of the oral mechanism. When they were 29 and 36 months old, a developmental oropharyngeal motor evaluation (Robbins & Klee, 1987) was added in an attempt to quantify oral motor abilities. At 29 months, 16 children did not cooperate with the Robbins and Klee evaluation. The mean structural score was 23.46 (SD = 0.99, SE = 0.14), and the mean functional score was 83.50 (SD = 19.98, SE = 2.88) for the 48 children who completed the exam. At 36 months, we were unable to complete the measure on 27 of the children. The mean structural score was 23.08 (SD = 2.10, SE = 0.35), and the mean functional score was 97.05 (SD = 11.04, SE = 1.82) for the 37 children who completed the tasks at Time 4. These scores are consistent with those reported for similar aged typically developing children by Robbins and Klee.

The hearing screenings confirmed normal hearing in the group as a whole, although there was some variation. At 17 months of age, 2 children refused to participate in the hearing test, and all of the others passed. At 21 months, a different 2 children refused to take the test, and all of the rest passed. When they were 29 months old, 9 children refused to participate, 3 failed the screening, and 52 passed. At 36 months, 11 children refused to participate, 1 failed, and 52 passed. No child failed at more than one visit. In addition, refusals to participate generally occurred at only a single visit, except for 4 children. One of those refused to be tested at all visits, another refused at all visits except the first, and 2 refused to be tested at two nonconsecutive test times. When they were 4 years old, the children returned to the laboratory for another battery of tests, but hearing screenings were not performed. Parents were asked if they had any concerns about hearing and about the presence of ear infections. No concerns or presently active otitis media were reported.

Scores on three language tests confirmed normal language development. These included the Clinical Evaluation of Language Fundamentals—Preschool (CELF–P; Wiig, Secord, & Semel, 1992), the Peabody Picture Vocabulary Test (PPVT–III; Dunn & Dunn, 1997), and the Expressive Vocabulary Test (EVT; Williams, 1997). The mean standard score on the CELF–P was 111.80 (range: 83–139). Similar scores were found for receptive vocabulary (PPVT–III, M = 112.61, range: 86–136) and expressive vocabulary (EVT,[1] M = 111.10, range: 90–129). Normal cognitive development was verified with the Kaufman Assessment Battery for Children (K–ABC; Kaufman & Kaufman, 1983). The mean Mental Processing Score on the K–ABC was 110.81 (range: 91–140).

Procedure. The NRT (Dollaghan & Campbell, 1998) was administered to the children as part of a battery that included the standardized tests of language and cognition identified above and a number of experimental measures. The children participated in two to three testing sessions, each lasting about 2 hr, typically completed within 1 week. The tests were usually administered in the same order, but the order was changed if necessary to maintain the children's interest. The CELF–P and EVT were generally administered during the first day of testing, the PPVT–III, K–ABC, and the NRT on the second. The standardized tests of language and cognition were administered according to the protocols in the published manuals. The entire NRT, including instructions, was administered through a Marantz tape recorder (PMD 221) set at Level 7, a comfortable listening level, following the procedure described by Dollaghan and Campbell (1998) and using taped stimuli that they generously provided. The children were typically seated at a table, and the tape recorder was placed in front of them. If the child was more comfortable working on the floor, however, then the test was administered with the child seated on the floor. The examiner pressed the play button on the tape recorder and the children heard the following instructions: "Now I will say some made up words. Say them after me exactly the way that I say them." Dollaghan and Campbell's procedure was adapted to allow the examiner to stop the tape after each nonword was presented to accommodate the children's young age, attention span, and the potential need for increased response time. Each child's responses were transcribed online using broad phonetic transcription. If the examiner was unable to transcribe all responses online, she or he completed the transcription later using both the videotaped and audiotaped recordings.

In general, the 4-year-old children in our study were able to complete the NRT without difficulty. Their cooperation was probably facilitated by their previous testing experience (as a part of our longitudinal study); a relaxed, child-directed, 10-min warm-up period before each testing session; and the fact that our testers had substantial training and experience in the administration of tests to preschool-age children. Testers gave generous positive feedback (e.g., enthusiastically saying "Good job!") after virtually every test item. If the tester were concerned about the child's attention, he or she usually said "Listen!" or "Ready?" before releasing the pause button and playing the next nonword. As an aid in transitioning from the previous activity to the NRT, testers introduced the NRT by saying: "We're going to play a talking game. The lady on the tape is going to say a funny word, then you say it just like the lady did." This introduction was an addition to the official test instructions, which were given by the voice on the cassette tape itself, as noted above.

The tapes of 14 randomly selected participants (22%) were reviewed to obtain an estimate of the kinds and amounts of difficulties encountered when administering the NRT to 4-year-old children. Responses to 5 nonwords (2.2% of the subsample) came after a prompt from the tester. Testers elected to repeat 18 nonwords (8.0%, or an average of about 1 nonword per participant) by rewinding the tape at the end of the test and readministering the item. Usually this occurred because it was clear to the tester that the child's attention had been distracted so that the child had not attended to the nonword the first time. In some cases, a repetition was necessary because a problem with operation of the pause button cut off or distorted all or part of a nonword. The testers requested that the child say the nonword a second time (without the benefit of hearing the taped stimulus again) on 3 nonwords (1.3%). This was done when a child spoke very softly or was unclear. When this occurred, the tester attempted to score the first production using the video and cassette recordings offline and referred to the second production only if it was impossible to score the first production.

If the child produced a nonword two or more times without being prompted to do so, and the tester believed the child's intention was to correct an error, then the final production was transcribed and scored. If the child repeated the nonword for any other reason (e.g., playfully), then the first production was transcribed and scored.

[1]One child did not complete the EVT.

Data reduction and analysis. The binary scoring system used by Dollaghan and Campbell (1998) was used, with a score of 1 for a correct repetition of a phoneme and a score of 0 if a phoneme was not correctly imitated. Substitutions and omissions were given a score of 0. Distortions of phonemes were scored as correct. The children's performance on the NRT was calculated as the percentage of phonemes produced correctly for each of the four syllable lengths of the nonwords and the TPPC summed across all four nonword lengths.

Although the "Late Eight" phonemes (Shriberg & Kwiatkowski, 1994) were not included in the NRT to ensure developmental appropriateness at 6 years of age, the nonwords do contain some phonemes that many typically developing 4-year-old children may not be expected to produce. Thus, in addition to calculating a score exactly as was done by Dollaghan and Campbell (1998), we calculated two additional scores: (a) one in which all affricates (/ʧ/ and /ʤ/) were excluded and (b) another in which all affricates and fricatives (/ʧ/, /ʤ/, /f/ and /v/) were excluded.

Responses to 10% of each sample were scored independently by a second trained listener, using the videotapes and audiotapes, to establish reliability. Phoneme-to-phoneme percentage of agreement was 99% for one- and two-syllable words, 92% for three-syllable words, 87% for four-syllable words, and 95% for the test as a whole.

Results and Discussion

Descriptive statistics on the percentage of consonants repeated correctly at each syllable level and the TPPC on the NRT were determined for each scoring method (see Table 1). A one-way analysis of variance (ANOVA) with scoring method as the factor was used to test for differences in TPPC across the methods. This was followed by a multivariate analysis of variance (MANOVA) and ANOVAs comparing the scoring methods across the four syllable levels. Correlations were used to determine the strength of the relationship between the scoring methods.

The ANOVA into which the TPPC from the three scoring methods were entered did not reach significance, $F(2, 189) = 1.08$, $p = .34$. A MANOVA in which each of the three scoring methods was compared for each syllable length was significant, Wilks's $\lambda = 0.911$, $p < .02$, $\eta_p^2 = .05$. One-way ANOVAs indicated a significant main effect for one- and two-syllable words, $F(2, 189) = 8.315$, $p < .0001$, $\eta_p^2 = .08$, and $F(2, 189) = 3.136$, $p < .05$, $\eta_p^2 = .03$, respectively. Post hoc Tukey tests ($p < .05$) indicated that scores for one-syllable nonwords were higher in the condition in which both affricates and fricatives were excluded than in both of the other conditions. No significant pairwise differences were found for two-syllable nonwords. Because the proportion of affricate or fricative phonemes is higher in one-syllable (25%) and two-syllable (30%) nonwords than in three-syllable (21%) and four-syllable (19%) nonwords, this result is in the expected direction. However, the effect size is extremely small. Correlations between the three scoring methods were high and significant for all syllable levels (one syllable: $r = .92–.97$; two syllables: $r = .84–.97$; three syllables: $r = .95–.98$; four syllables: $r = .98–.99$, $p < .0001$; TPPC: $r = .95–.99$, $p < .0001$), suggesting little to no practical difference between the scores. Thus, both the total score (TPPC) and the more fine-tuned analyses carried out at each syllable length support the claim that the original scoring method used by Dollaghan and Campbell (1998) for children age 6 years and older is appropriate for children who are as young as 4 years. These results are compatible with studies in which the CNRep has been used successfully with 4-year-old children (Adams & Gathercole, 2000; Bowey, 2001; Gathercole, 1995; Gathercole & Adams, 1994; Gathercole, Adams, & Hitch, 1994; Gathercole, Frankish, et al., 1999; Gathercole, Willis, et al., 1994) and suggest that the NRT is also appropriate for use with 4-year-old children.

When the normal 4-year-old children examined in this study are compared to the 6-year-olds described by Dollaghan and Campbell (1998, see Table 2), we observe that scores for the 4-year-old children are lower than those for the typically developing 6-year-olds at all

Table 1. Means, standard deviations, and ranges of scores on the Nonword Repetition Test for each syllable length and total percentage phonemes correct.

	Dollaghan & Campbell (1998) method		Affricates excluded		Fricatives and affricates excluded	
Syllable length	*M*	*SD* (Range)	*M*	*SD* (Range)	*M*	*SD* (Range)
One-syllable nonwords	80.47	11.13(50.00–100.00)	79.97	11.01(45.45–100.00)	87.15	11.24(55.56–100.00)
Two-syllable nonwords	87.58	11.41(55.00–100.00)	88.19	11.26(55.56–100.00)	91.96	9.39(64.00–100.00)
Three-syllable nonwords	79.46	13.87(50.00–100.00)	79.46	13.98(44.00–100.00)	81.25	13.42(45.45–100.00)
Four-syllable nonwords	67.27	21.14(16.67–97.22)	66.55	21.44(15.63–96.88)	69.53	21.62(17.86–100.00)
Total percentage phonemes correct	76.71	13.52(40.63–95.83)	76.43	13.23(41.90–95.40)	79.54	12.88(45.21–97.26)

Table 2. Nonword Repetition Test scores for 4-year-old children with normal language development, as reported in this study, and 6-year-old children without language impairment (LN) and with language impairment (LI), as reported by Dollaghan and Campbell (1998).

	LN 4-year-olds (*n* = 64)		LN 6-year-olds (*n* = 20)		LI 6-year-olds (*n* = 20)	
Syllable length	*M*	*SD*	*M*	*SD*	*M*	*SD*
One-syllable nonwords	80	11	91	6	86	9
Two-syllable nonwords	88	11	92	7	83	10
Three-syllable nonwords	79	14	90	9	68	20
Four-syllable nonwords	67	21	71	11	50	16
Total percentage phonemes correct	77	13	84	7	66	12

syllable levels and TPPC. However, the pattern of scores across syllable lengths is the same for both groups of children (higher on two-syllable nonwords than on one-syllable nonwords and then successively lower on three- and four-syllable nonwords). In addition, the scores of the 4-year-old children in this study are higher than those achieved by the 6-year-old children with language impairment at the three-syllable, four-syllable, and TPPC levels, the levels that differentiated children with language impairment (LI) from children with normal language (LN) in the Dollaghan and Campbell study. Thus, there appears to be a developmental difference for typically developing children on the NRT, but that difference is not as large as that between 6-year-old children who are LN and LI. A study that directly compares 4- to 6-year old children is necessary to verify this observation.

It is not clear why the 4-year-old children in this study did so much better with two-syllable nonwords than with one-syllable nonwords. One possibility is that the combination of the higher percentage of fricatives combined with less information (e.g., only one syllable, lack of trochaic structure) made the one-syllable words more difficult. The lack of any real difference between one- and two-syllable words for the normal children studied by Dollaghan and Campbell (1998), combined with the poorer performance of the 4-year-olds in this study, suggests that something about the one-syllable words is more difficult. Alternatively, something about the two-syllable words may be easier. Research in which the effects of phonological structure and syllable length of one- and two-syllable nonwords on repetition accuracy of young children is systematically explored is necessary to clarify this finding.

Study 2

Method

In Study 2, we compared scores of children who had been delayed in language production at 16 months of age to those of children who had never shown any language delay to determine whether the NRT scores discriminated between the two groups.

Participants. The same 64 children who participated in Study 1 also served in Study 2. Although all of those children were clearly within the normal range in language development at the time of this study, 20 of them (11 males and 9 females) had been delayed at the early stages of language development. When they were 16 months of age, they had scored at or below the 10th percentile (using gender-specific norms) for vocabulary production on the MacArthur–Bates Communicative Development Inventory: Words and Sentences (CDI:WS; Fenson et al., 1993). Use of this parent report instrument allowed us to obtain a more detailed sample of the children's vocabulary on which to base our judgment of group classification, a practice that is important given the wide variability and limited stability of individual differences in language development at these early ages. Fenson et al. (2000) demonstrated that the CDI provides an authentic reflection of behaviorally measured individual differences in early language development. In addition, Bates et al. (2002) and Rodrigue, Jeanette, Shen, and Thal (2002) demonstrated that the rate of growth in vocabulary size from 16 to 29 months of age was virtually identical when measured from a spontaneous language sample and the CDI:WS in this particular cohort of children. Thus, we are confident that the CDI:WS has provided an accurate measure of vocabulary size in the children used for this study.

Sixteen of the children with language delay scored above the 10th percentile on vocabulary comprehension at 16 months of age on the CDI: Words and Gestures (CDI:WG), and 4 scored below the 10th percentile in vocabulary comprehension. Thus, they reflect the variability seen in the population described as specifically language impaired in the literature, although their language abilities are in the normal range at 4 years of age. The other 44 children had scored above the 10th percentile on both comprehension and production at 16 months of age. Specifically, vocabulary production scores on the CDI:WS showed that 10 children were ranked between the 12th and 25th percentiles, 15 between the 26th and 50th percentiles, 18 between the 51st and 75th percentiles, and 1 at the 96th percentile. Vocabulary comprehension scores on the CDI:WG showed that 7 children were ranked between the 13th and 25th percentiles, 12 between the 26th and 50th percentiles, 23 between the 51st and 75th percentiles, 1 at the 79th percentile, and 1 at the 82nd percentile. No differences between groups in oral motor development or hearing/otitis media history were found.

Despite the lack of clinical delay when they were 4 years old, the children with a history of language delay (henceforth *HLD*) scored significantly lower than those with a history of typical language development (henceforth *HTD*) on the tests of language and cognitive processing, CELF–P Receptive Language, $F(1, 62) = 6.35$, $p < .02$, $\eta_p^2 = .09$; CELF–P Expressive Language, $F(1, 62) = 4.09$, $p < .05$, $\eta_p^2 = .06$; PPVT–III, $F(1, 62) = 3.99$, $p < .05$, $\eta_p^2 = .06$; EVT, $F(1, 62) = 7.57$, $p < .008$, $\eta_p^2 = .11$; and K–ABC Mental Processing Score, $F(1, 62) = 6.76$, $p < .01$, $\eta_p^2 = .10$. Thus, the relative position of the HLD as a group with regard to HTD appears to have remained unchanged from the earlier measurement of language at 16 months of age (see Table 3), although the effect sizes are all small.

Procedure. In this study, we were interested in whether there were differences between the HTD children and the HLD children on the NRT TPPC and whether TPPC alone was as good a predictor of group membership as measures of language and mental processing (or a combination of the two). Because the results of Study 1 showed no important differences between scoring methods, only the scoring method originally used by Dollaghan and Campbell (1998) was used for the NRT.

Table 3. Comparison of 4-year-old children with a history of typical language development (HTD) to those with a history of language delay (HLD) on standardized tests of language and mental processing.

	HTD (*n* = 44)		HLD (*n* = 20)	
Syllable length	*M*	*SD*	*M*	*SD*
CELF–P Receptive Language	114.73	12.17	106.35	12.68
CELF–P Expressive Language	112.84	10.56	106.30	14.81
PPVT–III	114.19	10.55	109.45	9.82
EVT	113.00	8.31	107.00	10.82
K–ABC MPS	113.14	10.96	106.40	9.25

Note. CELF–P = Clinical Evaluation of Language Fundamentals—Preschool; PPVT–III = Peabody Picture Vocabulary Test—Third Edition; EVT = Expressive Vocabulary Test; K–ABC MPS = Kaufman Assessment Battery for Children, Mental Processing Score.

Data reduction and analysis. A *t* test was used to compare the groups on the NRT TPPC. A mixed-model, repeated measures ANOVA was used to compare the groups at each of the four nonword syllable lengths. Logistic regression was used to determine the sensitivity and specificity of the language, sequential processing, and TPPC measures for differentiating the two groups. This was followed by likelihood ratio analyses to evaluate the ability of TPPC scores alone to accurately identify the two groups of children.

Results and Discussion

Given the results from the K–ABC Mental Processing Score, one important question is whether HLD children are simply less skilled cognitively than HTD children. To answer that question, we used a MANOVA with group as the fixed factor and standard scores from the sequential and simultaneous processing subtests of the K–ABC as the dependent variables. The MANOVA was significant, Wilks's $\lambda = .905$, $p < .05$, $\eta_p^2 = .09$. One-way analyses of the two different subtests indicated a significant difference between the groups for sequential processing, $F(1, 62) = 5.91$, $p < .02$, $\eta_p^2 = .09$, but not for simultaneous processing, $F(1, 62) = 2.09$, *ns*. Because of the significant ANOVA, the sequential processing subtests (Hand Movements, Number Recall, Word Order) were also examined via MANOVA. This analysis did not reach significance (Wilks's $\lambda = .905$, $p < .11$). Thus, the two groups are matched on the simultaneous processing measures that include gestalt closure, face recognition, pattern design, and temporal–spatial ability, and these have substantial overlap with tasks more commonly used to match children on performance IQ. In other words, TLD children and HTD children were not different on all mental processing abilities; they were different only on the summary subtest score of analytic processing tasks that are thought to be mediated in the left hemisphere (Kaufman & Kaufman, 1983).

Group performance on nonword repetition. Next, the groups were compared on TPPC using a *t* test. This was followed by a 2 (Group) × 4 (Length) mixed-model repeated measures ANOVA in which the groups were compared on the percentage of phonemes produced correctly at each of the four syllable levels. The mean scores at each syllable level and TPPC are displayed in Table 4. The results indicated a significant difference between the two groups on TPPC, $t(62) = 3.60$, $p < .001$, Cohen's $d = .16$. The ANOVA indicated significant main effects of group, $F(1, 62) = 12.45$, $p < .001$, $\eta_p^2 = .17$, and length, $F(1, 62) = 45.60$, $p < .0001$, $\eta_p^2 = .42$, and a significant Group × Length interaction, $F(1, 62) = 4.13$, $p < .02$, $\eta_p^2 = .06$. Post hoc Scheffé tests with alpha set at <.01 revealed significant differences between the groups at the three- and four-syllable levels. Thus, it is clearly

Table 4. Comparison of 4-year-old children with a history of typical language development (HTD) to those with a history of language delay (HLD) on the Nonword Repetition Test.

Test item	HTD (n = 44)		HLD (n = 20)	
	M	SD	M	SD
PPC 1-SYLL	82.39	10.67	76.25	11.24
PPC 2-SYLL	89.32	11.44	83.75	10.62
PPC 3-SYLL	83.20	11.15	71.25*	15.91
PPC 4-SYLL	72.79	17.58	55.14*	23.59
TPPC	80.47	10.88	68.44*	15.29

Note. PPC = percent phonemes correct, SYLL = syllable, TTPC = total percent phonemes correct.

*$p < .01$.

the case that children who had been delayed in early language development performed significantly more poorly on the NRT than did children with a history of typical development. The effect sizes for the three- and four-syllable nonwords, and TPPC, although small, are larger than those for the CELF–P, PPVT–III, EVT, and K–ABC. These findings for children with a history of early language delay replicate those reported for children with SLI by Dollaghan and Campbell (1998) and by Ellis Weismer et al. (2000). The closeness of the scores of each of the groups of 4-year-old children examined in this study to the comparable group in those studies (HTD/ normal language and HLD/SLI) is remarkable given the significantly higher ages of the children studied by Dollaghan and Campbell and by Ellis Weismer et al. (6–9 years and 7–9 years old, respectively).

Adequacy of classification. All of the language variables, TPPC, and the sequential processing variables from the K–ABC were entered into a logistic regression to determine whether the two groups could be predicted accurately. The model estimation terminated after five iterations, and the omnibus test was significant, $\chi^2(12, N = 64) = 25.256, p < .02$. The coefficients for only two variables, TPPC and the EVT, were significantly different from zero (Wald = 5.18, $p < .02$, and 4.39, $p < .04$, respectively). Overall, 79.4% of the 64 children—86% of the HTD and 65% of the HLD—were correctly classified. When the EVT was removed from the analysis, only 55% of the HLD children were accurately identified. When TPPC was removed from the model, and all of the language measures were added, the omnibus test only approached significance, $\chi^2(8, N = 64) = 14.033, p < .08$, and specificity was reduced to 40% accurate identification of the HLD children. On the other hand, when TPPC was the only factor in the model, the omnibus test was significant, $\chi^2(1, N = 64) = 11.309, p < .001$, and 40% of the HLD children were accurately identified. Taken together, these results indicate that the NRT has impressive ability to identify children with weaker language abilities at 4 years of age, a finding that is consistent with reports for the CNRep with typically developing children (Adams & Gathercole, 2000; Avons et al., 1998; Gathercole & Adams, 1994; Gathercole, Willis, Emslie, & Baddeley, 1992) and SLI (Botting & Conti-Ramsden, 2001), and for the NRT for children with SLI (Dollaghan & Campbell, 1998). However, TPPC alone was not as sensitive as when it was combined with a test of expressive vocabulary (the EVT). This finding is compatible with arguments made by Ellis Weismer et al. (2000).

Following Dollaghan and Campbell (1998) and Ellis Weismer et al. (2000), we also calculated likelihood ratios (Sackett, Haynes, Guyatt, & Jugwell, 1991) to determine the power of the NRT to rule in or rule out a history of early language delay in 4-year-old children who have language scores in the normal range. This could be particularly useful for identifying children who need more careful watching and more regular follow-up as they progress through school. In the first set of analyses, we used the criteria established by Dollaghan and Campbell and used by Ellis Weismer et al. to make a direct comparison to those two studies. A TPPC score of 70% or lower was defined as a positive result (to rule in history of early language delay) and a score of 81% or higher was defined as a negative result (ruling out a history of early language delay). To determine the likelihood (LH) ratio for a positive result, the proportion of HLD children with scores at or below 70% (true positives) was divided by the proportion of HTD children with scores at or below 70%. The LH ratio for a negative result was determined by dividing the proportion of HLD children with scores at or above 81% by the proportion of HTD children at or above 81% (true negatives). Likelihood ratios for intermediate high scores (71%–74% and 75%–80%) also were calculated. In the second set of analyses, we used the more extreme cutoff scores established by Ellis Weismer et al. (at or below 60% for positive, at or above 90% for negative, and 61%–89% for intermediate) to maximize our ability to rule in or out the history of early language delay and to allow a direct comparison with that study. Likelihood ratios greater than 1 indicate that TPPC scores in the range tested are more likely to come from a child with a history of early language delay; LH ratios of less than 1 indicate that the scores are more likely to come from a child with a history of no language delay at 16 months. These results are displayed in Table 5.

Likelihood ratio analyses in which the Dollaghan and Campbell (1998) cutoffs were used are presented in the upper panel of Table 5. The LH ratio for accurately classifying a 4-year-old child as having a history of language delay at 16 months of age was 3.33 (0.60/0.18), indicating that TPPC scores of 70% or lower were more

Table 5. Likelihood ratios for total percentage phonemes correct (TPPC) on the Nonword Repetition Test, based on history of language delay at 16-months of age.

	HLD		HTD		
TPPC	Number	Proportion	Number	Proportion	Likelihood ratio
Dollaghan & Campbell (1998) cut points					
≤70	12	.60	8	.18	3.33
71–74	2	.10	5	.11	0.91
75–80	1	.05	6	.14	0.36
≥81	5	.25	25	.57	0.44
Ellis Weismer et al. (2000) extreme cut points					
≤60	6	.30	3	.07	4.28
61–89	11	.60	31	.70	0.86
≥90	3	.15	10	.23	0.65

than three times more likely to come from HLD children than from HTD children. The LH ratio for accurately ruling out a history of language delay at 16 months in a 4-year-old child was 0.44 (0.25/0.57), indicating that TPPC scores of 81% or greater are less than five-tenths as likely to come from HLD children as from HTD children. The lower panel of Table 5 displays the LH ratios with the same extreme scores used by Ellis Weismer et al. (2000). In this analysis, the LH ratio for accurately classifying a 4-year-old child as having a history of language delay at 16 months of age was 4.28 (0.30/0.07) and that for accurately ruling out a history of language delay at 16 months in a 4-year-old child was 0.65 (0.15/0.23). That is, a TPPC score of 60% or lower was more than four times more likely to come from HLD children than from HTD children, and a TPPC score of 90% or higher was less than seven-tenths as likely to come from HLD children as from HTD children.

These results are remarkably like those reported for older language-impaired children by Ellis Weismer et al. (2000). In that study, as in this, the most discriminating ratios for ruling in a history of early language delay were in the "intermediate high" range (cf. Sackett et al., 1991), suggesting that poor performance on the NRT can help identify children with a history of early language delay but that corroboration from additional factors will be needed to make a stronger classification. Similarly, the results for ruling out a history of early language delay would be considered "intermediate low," indicating that a clear classification is not possible with the NRT alone. These results are also consistent with the categorical regression analysis reported above. It is important to remember that the HLD children were no longer language delayed at 4 years of age. A study with children who are language delayed at age 4 may demonstrate that the NRT has stronger independent discriminant power than is indicated in this study.

Study 3
Method

In the third study, we examined the relation between NRT scores and scores on specific subtests of language and mental processing with the two groups of children recombined into the single, larger group. We chose to recombine the groups for this study because the HLD group was very small for the analysis we wanted to carry out, language was within the normal range for all of the children at the time that the NRT was administered, and Study 2 demonstrated that the NRT score alone was not sufficient to differentiate HTD children from HLD chilren. This was an exploratory study designed to see whether the strong relations between language and the CNRep reported in the literature would also be found for the NRT.

Participants. The same 64 children who participated in Studies 1 and 2 also served in Study 3. One child was omitted because the child did not have scores for the EVT, leaving a final sample size of 63 children.

Procedure. As in Study 2, the Dollaghan and Campbell (1998) scoring method was used for the NRT. Because our interest was in relations between scores on the NRT and specific linguistic and nonlinguistic cognitive skills, we used standard scores from subtests of the CELF–P and K–ABC as well as standard scores from the PPVT–III and EVT. The PPVT–III, EVT, and subtests from the CELF–P provided a variety of measures of language knowledge (subtest details are provided in the Appendix). Although scores on these measures are all significantly intercorrelated (at R = .39 to R = .62, p < .05, in this sample of children), there are conceptual, structural, and response requirement differences between the individual tasks that may relate differently to performance on the NRT. We hypothesized that there would be significant relations between the NRT TPPC and all of the language measures.

The K–ABC was chosen to examine intellectual ability because it is a process-oriented test that contains a number of scales that tap various content and sensory domains (subtest details are provided in the Appendix). All of the tasks that comprise the Sequential Processing scale require information to be arranged in serial order and the use of short-term memory for resolution. The tasks in the Simultaneous Processing scale require simultaneous integration and synthesis of input, and they focus on spatial, analogic, or organizational problems. The Simultaneous Processing subtests resemble (in fact, some are exact replicas of) measures used on tests of performance IQ. Although the Sequential and Simultaneous Processing scales are considered to be equally important to intellectual functioning, they have been shown to be reasonably distinct through factor and discriminant analysis (Kaufman & Kaufman, 1983), and have potential to further our understanding of the nature of the NRT task. In particular, we hypothesized significant relations between the NRT TPPC score and sequential processing measures. In addition, a finding of meaningful relations between TPPC and some subtests of the Sequential Processing scale and not others, and between some of the language subtests but not others, may help to clarify the nature of the mental processes tapped by the NRT (Dollaghan & Campbell, 1998).

Data reduction and analysis. First, standard scores from the EVT, PPVT–III, and each of the subtests of the CELF–P and the K–ABC were entered into an exploratory principal-components analysis with varimax rotation. Principal-components analysis identifies variables within a larger set that are more highly correlated with each other, and organizes them into a smaller number of separate factors that are regarded as reflecting underlying processes that explain the relations between the variables (Tabachnick & Fidell, 2001). After initial extraction, a procedure called *rotation* is applied to help interpret the factors. The varimax rotation is a commonly used technique for creating components that are not correlated with each other (i.e., are orthogonal) and thus may be considered to be independent of each other.

Results and Discussion

The principal-components analysis yielded four factors with eigenvalues greater than 1 that, together, accounted for 61.27% of the variance in the data. However, 4 of the 16 variables had loadings greater than .40 on two or more factors. Under such conditions, a maximum likelihood method of extraction is recommended because it calculates factor loadings that maximize the likelihood of observing the factors that characterize the observed correlation matrix (Tabachnick & Fidell, 2001). The maximum likelihood analysis with varimax rotation also yielded four factors, accounting for 50.95% of the variance, and only two of them had loadings greater than .40 on two factors. The goodness-of-fit test indicated a good fit of the model to the data, $\chi^2(62, N = \text{xx}) = 45.24$, $p < .94$.[2] The factors identified with this model are summarized with the factor loadings after rotation in Table 6 (loadings of .400 and above are in boldface type).

The first factor (accounting for 17.30% of the variance, eigenvalue = 2.77) was defined by the Word Structure and Sentence Structure subtests of the CELF–P and by the EVT. The Word Order and Triangles subtests of the K–ABC and the Formulating Labels and Linguistic Concepts subtests of the CELF–P also loaded on this factor. The last two subtests loaded more strongly on the second factor, however. The defining feature of Factor 1 appears to be expressive language.

The second factor (accounting for 15.81% of the variance, eigenvalue = 2.53) was defined by the Basic Concepts and Sentence Structure subtests of the CELF–P. The Formulating Labels and Linguistic Concepts subtests of the CELF–P, and the PPVT–III, also made substantial contributions to this factor. The defining feature of this factor appears to be language comprehension. These language factors accounted for a cumulative 33.11% of the variance (out of a total 51.91%) in the data.

The third factor (accounting for 9.63% of the variance, eigenvalue = 1.54) was defined by the NRT TPPC with the Number Recall subtest of the K–ABC also loading on the factor. This appears to reflect verbal sequences that have little semantic or morphosyntactic content. A relation between nonword repetition and digit span is consistent with all of the earlier studies in which the CNRep has been used. However, the finding that nonword repetition (as measured by the NRT) and digit span (as measured by the K–ABC Number Recall subtest) did not form a factor with other language measures is inconsistent with earlier claims of the specific linguistic (in particular, phonological) nature of the mechanism tapped by nonword repetition.

The fourth factor (accounting for 8.21% of the variance, eigenvalue = 1.31) was defined solely by the Magic Window subtest of the K–ABC. The appearance of this factor highlights the very different nature of this visual gating task. The loadings for Gestalt Closure and Face Recognition were very low on all of the factors. The same is true for Hand Movement, although, if .30 is used as a cutoff, it may be considered as making a small contribution to Factor 3.

The most surprising result of this study was Factor 3. Because the NRT had differentiated children with high and low language skills in Study 2, we anticipated that TPPC would be represented in a factor that also included language variables. This expectation was strengthened

[2] A nonsignificant chi-square indicates a good fit.

Table 6. Factor loadings after varimax rotation for scores on the Nonword Repetition Test (NRT), and on language comprehension, language production, sequential mental processing, and simultaneous mental processing tests.

Measure	Factors 1	2	3	4
NRT TPPC	.287	.167	**.942**	.003
Language comprehension				
PPVT–III	.309	**.502**	.149	.116
CELF–P				
Linguistic Concepts	**.452**	**.578**	.274	.003
Basic Concepts	.221	**.818**	.009	.138
Sentence Structure	.181	**.604**	.138	.276
Language production				
EVT	**.695**	.368	.004	.009
CELF–P				
Recalling Sentences	**.604**	.360	.152	.007
Formulating Labels	**.431**	**.591**	.190	−.007
Word Structure	**.712**	.117	.229	.184
Mental Processing				
K–ABC Sequential				
Hand Movements	.003	.108	.301	.183
Number Recall	.339	.171	**.430**	−.002
Word Order	**.539**	.345	.236	−.001
K–ABC Simultaneous				
Gestalt Closure	.296	.192	.117	.223
Triangles	**.491**	.184	.180	.214
Face Recognition	.005	.003	.165	.278
Magic Window	.157	.148	−.009	**.971**

Note. Factor loadings of .400 and above are in boldface type.

by the results of a recent study of 633 children who were between the ages of 4 and 6 years (Alloway et al., 2004). An exploratory principal-components analysis in that study yielded two factors, one of which contained digit recall, word recall, CNRep scores, and two sentence repetition tasks. If the NRT and the CNRep both tap the same underlying processing mechanism, then the NRT TPPC scores should have loaded onto a factor that also contained the Repeating Sentences subtest of the CELF–P, at a minimum. Instead, we found that TPPC appeared to be independent of the tests of expressive and receptive language. Even the Word Order subtest of the K–ABC did not load on Factor 3. This could potentially be explained by the different response requirement (the K–ABC Word Order subtest requires the child to point to pictures of the words in the order in which he or she heard them rather than to repeat them verbally). However, Adams and Gathercole (2000) demonstrated correlations between the CNRep and a similar task and argued that relations between indexes of language and working memory were not due to the common output requirements of the tasks.

One potential explanation for our results is the difference in structure between the NRT and the CNRep. As noted in the beginning of this article, the NRT is less confounded with children's knowledge of linguistic forms than is the CNRep and, therefore, may be considered a purer measure of underlying short-term memory process. It is possible that the phonological specificity of the CNRep is really a reflection of the children's level of linguistic experience. This argument is also consistent with the fact that the word span task that we used (K–ABC Word Order) loaded on a language factor rather than with the NRT. The finding that NRT did not load with other variables in our factor analysis should, however, be viewed with caution given that performance on number recall (i.e., digit span) also did not load with language measures as one would have expected, based on previous findings. Nevertheless, differences in findings with CNRep compared with NRT with regard to the relationship between working memory and language development warrant a direct comparison of these two tests.

In addition to direct comparisons of the two most widely used nonword repetition tests, exploration of a number of additional strands of research would be helpful to determine whether nonword repetition abilities reflect the working of a specific phonological processor or some more general mechanism. It is notable that virtually all of the work in which this aspect of the model has been explored to date has focused on children who are age 4 years or older, children who already have large receptive and expressive vocabularies. Explorations with younger children who are truly involved in the process of figuring out words and phonological representations may be a better place to look to answer this developmental question. Evans and MacWhinney (1999) demonstrated that children with both receptive and expressive SLI used different strategies for determining sentence subjects than children with purely expressive SLI and that both used strategies that differed from those used by children with normal language development. These and other researchers have used models that come from dynamic systems and connectionist perspectives (cf. Elman, 2003; Smith & Samuelson, 2003; Thelen & Bates, 2003) to design their studies and interpret their results. These kinds of models may be of particular value for clinical purposes in that they can help researchers understand what kinds of strategies children with poorer working memory use to develop normal language skills and how those strategies are learned and changed.

Examining children at younger ages will not be a simple task. However, Roy and Chiat (2004) recently developed a repetition test for use with children as young as 2 years that includes both words and nonwords, and controls for prosody. This is a good beginning and appears to be likely to meet their goal of identifying very young children with word and nonword repetition weaknesses. Further adaptation to control for things such as presence of consonant clusters, syllables that correspond to English words, later developing consonants, and frequency of occurrence in real words may make such a test a useful tool for comparison to nonverbal tasks and separating results affected by linguistic experience from those that can be explained by more general processing in the period during which the problem of representing information through words is a major focus of children.

General Discussion and Conclusions

We have reported three studies of NRT scores (Dollaghan & Campbell, 1998) in 4-year-old children with language that is in the normal range on a number of receptive and expressive tests of language. The results of the first study were straightforward: Four-year-old children had no difficulty completing the test, and the original scoring system was appropriate. Thus, researchers and clinicians should have no hesitation in using the NRT with 4-year-old children.

In Study 2, we divided the 4-year-old children into two groups: (a) one that had a history of language delay at 16 months of age and (b) another with no history of delay. The results supported our hypothesis that the HLD children would score significantly lower than HTD children on the NRT. This adds to the body of literature in which variation in nonword repetition ability is linked to variation in language skills in children with language impairment or a history of language impairment (Bishop et al., 1996; Dollaghan & Campbell, 1998; Edwards & Lahey, 1998; Ellis Weismer et al., 2000, Gathercole & Baddeley, 1990a, 1990b, 1995; Montgomery, 1995a, 1995b; Sahlen, Reuterskioeld-Wagner, Nettelbladt, & Radeborg, 1999) as well as those with normal histories (e.g., Adams & Gathercole, 1995, 2000; Baddeley, Gathercole, & Papagno, 1998; Gathercole & Adams, 1993; Gathercole & Baddeley, 1990b; Gathercole, Hitch, Service, & Martin 1997; Gathercole, Service, et al., 1999; Gathercole et al., 1992). Our results indicate that weaker nonword repetition skills (that are thought to index some aspect of working memory) characterize children with weaker language skills, a finding that is consistent with all of the earlier work. They also indicate that nonword repetition alone is not a sufficient index of weak language abilities, a finding consistent with that of Ellis Weismer et al. (2000) for older children with specific language impairment.

Study 3 was an exploratory study designed to tease apart the variables that are related to NRT performance. The results were quite unexpected and did not support our hypothesis that TPPC would load on a factor that also contained language variables from the CELF–P, EVT, and PPVT–III and sequential processing variables from the K–ABC. In the factor analyses, one factor was defined by TPPC and the Number Recall subtest of the K–ABC. The fact that nothing else loaded on this factor is quite remarkable in the context of earlier findings and may provide a challenge to claims that nonword repetition and digit span provide a direct measure of a domain-specific phonological short-term memory mechanism. We have speculated that these results may reflect differences between the measures used in other studies and the NRT and noted that, among tests of nonword repetition, the NRT is currently the least confounded by word-likeness. We also suggested caution in generalizing this specific result until it can be replicated, because it diverges dramatically from the large body of literature in which the CNRep was used. Finally, we suggested some alternative models for exploring this issue.

Acknowledgments

This research was supported by Grant RO1 DC000482 from the National Institute on Deafness and Other Communicative Disorders (principal investigator: Donna Thal). We thank Cassandra Callaghan and Julia Nance for their assistance with various aspects of data and article preparation and Julia Evans for stimulating discussions that helped shape the design of Study 2.

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.

Alloway, T., Gathercole, S., Willis, C., & Adams, A. (2004). A structural analysis of working memory and related skills in young children. *Journal of Experimental Child Psychology, 87,* 85–106.

Avons, S., Wragg, C. A., Cupples, L., & Lovegrove, W. J. (1998). Measures of phonological short-term memory and their relationship to vocabulary development. *Applied Psycholinguistics, 19,* 583–601.

Baddeley, A. (1992, January 31). Working memory. *Science, 255,* 556–559.

Baddeley, A. (2000). The episodic buffer: A new component of working memory? *Trends in Cognitive Science, 4,* 417–423.

Baddeley, A. (2003). Working memory and language: An overview. *Journal of Communication Disorders, 36,* 189–208.

Baddeley, A., Chincotta, D., Stafford, L., & Turk, D. (2002). Is the word length effect in STM entirely attributable to output delay? Evidence from serial recognition. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology, 55*(A), 353–369.

Baddeley, A., Gathercole, S., & Papagno, C. (1998). The phonological loop as a language learning device. *Psychological Review, 105,* 158–173.

Baddeley, A. D., & Hitch, G. J. (1974). Working memory. In G. Bower (Ed.), *The psychology of learning and motivation* (Vol. 8, pp. 47–90). New York: Academic Press.

Baddeley, A., & Hitch, G. (2000). Development of working memory: Should the Pascual-Leone and the Baddeley and Hitch models be merged? *Journal of Experimental Child Psychology, 77,* 128–137.

Baddeley, A. D., Logie, R. H., & Ellis, N. C. (1988). Characteristics of developmental dyslexia. *Cognition, 29,* 197–228.

Bates, E., Dale, P., Fenson, L., Goodman, J., Jackson-Maldonado, D., Marchman, V., & Thal, D. (2002, July). *The MacArthur CDI: Past, present, and future.* Symposium conducted at the joint meeting of the International Conference for the Study of Child Language and the Symposium on Research in Child Language Disorders, Madison, WI.

Bates, E., & Goodman, J. (1997). On the inseparability of grammar and the lexicon: Evidence from acquisition, aphasia, and real-time processing. *Language & Cognitive Processes, 12,* 507–584.

Bishop, D. V., 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.

Bowey, J. A. (2001). Nonword repetition and young children's receptive vocabulary: A longitudinal study. *Applied Psycholinguistics, 22,* 441–469.

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

Cocchini, G., Logie, R. H., Della Sala, S., MacPherson, S. E., & Baddeley, A. D. (2002). Concurrent performance of two memory tasks: Evidence for domain-specific working memory systems. *Memory & Cognition, 30,* 1086–1095.

Dollaghan, C., Biber, M., & Campbell, T. (1993). Constituent syllable effects in a nonsense-word repetition task. *Journal of Speech and Hearing Research, 36,* 1051–1054.

Dollaghan, C. A., Biber, M. E., & Campbell, T. F. (1995). Lexical influences on nonword repetition. *Applied Psycholinguistics, 16,* 211–222.

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. (1997). *Peabody Picture Vocabulary Test—III*. Circle Pines, MN: AGS.

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. E., 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.

Elman, J. (2003). Development: It's about time. *Developmental Science, 6,* 430–433.

Elman, J., Bates, E., Johnson, M., Karmiloff-Smith, A., Parisi, D., & Plunkett, K. (1996). *Rethinking innateness*. Cambridge, MA: MIT Press.

Evans, J., & MacWhinney, B. (1999). Sentence processing strategies in children with expressive and expressive–receptive specific language impairments. *International Journal of Language and Communication Disorders, 34,* 117–134.

Fenson, L., Bates, E., Dale, P., Goodman, J., Reznick, J. S., & Thal, D. (2000). Measuring variability in early child language: Don't shoot the messenger. *Child Development, 71,* 323–328.

Fenson, L., Dale, P., Reznick, J. S., Thal, D., Bates, E., Hartung, J., et al. (1993). *The MacArthur Communicative Development Inventories: Manual and norms.* San Diego, CA: Singular.

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

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

Gathercole, S. E., & Adams, A. M. (1994). Children's phonological working memory: Contributions of long-term knowledge and rehearsal. *Journal of Memory & Language, 33,* 672–688.

Gathercole, S., Adams, A. M., & Hitch, G. J. (1994). Do young children rehearse? An individual-differences analysis. *Memory & Cognition, 22,* 201–207.

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

Gathercole, S. E., & Baddeley, A. D. (1990b). The role of phonological memory in vocabulary acquisition: A study of young children learning new names. *British Journal of Psychology, 81,* 439–454.

Gathercole, S. E., & Baddeley, A. D. (1995). Short-term memory may yet be deficient in children with language impairments: A comment on van der Lely & Howard (1993). *Journal of Speech and Hearing Research, 38,* 463–472.

Gathercole, S., Frankish, C. R., Pickering, S. J., & Peaker, S. (1999). Phonotactic influences on short-term memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 25,* 84–95.

Gathercole, S. E., Hitch, G. J., Service, E., & Martin, A. J. (1997). Phonological short-term memory and new word learning in children. *Developmental Psychology, 33,* 966–979.

Gathercole, S. E., & Pickering, S. J. (1999). Estimating the capacity of phonological short-term memory. *International Journal of Psychology, 34,* 378–382.

Gathercole, S. E., & Pickering, S. J. (2000). Assessment of working memory in six- and seven-year-old children. *Journal of Educational Psychology, 92,* 377–390.

Gathercole, S. E., Service, E., Hitch, G. J., Adams, A.-M., & Martin, A. J. (1999). Phonological short-term memory and vocabulary development: Further evidence on the nature of the relationship. *Applied Cognitive Psychology, 13,* 65–77.

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,* 103–127.

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.

Jarrold, C., Baddeley, A. D., & Phillips, C. E. (2002). Verbal short-term memory in Down syndrome: A problem of memory, audition, or speech? *Journal of Speech, Language, and Hearing Research, 45,* 531–544.

Kamhi, A. G., & Catts, H. W. (1986). Toward an understanding of developmental language and reading disorders. *Journal of Speech and Hearing Disorders, 51,* 337–347.

Kamhi, A. G., Catts, H. W., Mauer, D., Apel, K., & Gentry, B. F. (1988). Phonological and spatial processing abilities in language- and reading-impaired children. *Journal of Speech and Hearing Disorders, 53,* 316–327.

Kaufman, A., & Kaufman, N. (1983). *Kaufman Assessment Battery for Children*. Circle Pines, MN: AGS.

Montgomery, J. W. (1995a). Examination of phonological working memory in specifically language-impaired children. *Applied Psycholinguistics, 16,* 355–378.

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

Robbins, J., & Klee, T. (1987). Clinical assessment of oropharyngeal motor development in young children. *Journal of Speech and Hearing Disorders, 52,* 271–277.

Rodrigue, S., Jeanette, B., Shen, J., & Thal, D. (2002, July). *Does language status at 16 months relate to later spontaneous language production?* Poster presented at the joint meeting of the International Conference for the Study of Child Language and the Symposium on Research in Child Language Disorders, Madison, WI.

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,* 223–234.

Sackett, D. L., Haynes, R. B., Guyatt, G. H., & Jugwell, P. (1991). *The interpretation of diagnostic data* (2nd ed.). Boston: Little, Brown.

Sahlen, B., Reuterskioeld-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.

Shriberg, L. D., & Kent, R. D. (1982). *Clinical phonetics*. New York: Wiley.

Shriberg, L. D., & Kwiatkowski, J. (1994). Developmental phonological disorders: A clinical profile. *Journal of Speech and Hearing Research, 37,* 1100–1126.

Simkin, Z., & Conti-Ramsden, G. (2001). Non-word repetition and grammatical morphology: Normative data for children in their final year of primary school. *International Journal of Language and Communication Disorders, 36,* 395–404.

Smith, L. B., & Samuelson, L. (2003). Different is good: Connectionism and dynamic systems theory are complementary emergentist approaches to development. *Developmental Science, 6,* 434–439.

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

Tabachnick, B., & Fidell, L. (2001). *Using multivariate statistics*. Needham Heights, MA: Allyn & Bacon.

Thelen, E., & Bates, E. (2003). Connectionism and dynamic systems: Are they really different? *Developmental Science, 6,* 378–391.

Wiig, S. E., Secord, W., & Semel, E. (1992). *Clinical Evaluation of Language Fundamentals—Preschool*. San Antonio, TX: The Psychological Corporation.

Williams, K. (1997). *Expressive Vocabulary Test*. Circle Pines, MN: AGS.

Received March 3, 2004

Accepted April 21, 2005

DOI: 10.1044/1092-4388(2005/103)

Contact author: Donna Thal, 402 Brighton Avenue, Cardiff, CA 92007. E-mail: dthal@mail.sdsu.edu

Appendix. Description of the language and intellectual processing tests and subtests used in Study 3.

Language Comprehension

Peabody Picture Vocabulary Test—Third Edition (PPVT–III; Dunn & Dunn, 1997). The child hears an open class word and points to a picture that represents the word, choosing from an array of four black-and-white line drawings.

Clinical Evaluation of Language Fundamentals—Preschool (CELF–P; Wiig, Secord, & Semel, 1992).

Linguistic Concepts. The child hears a sentence-length command and points to an animal or combination of animals in a color drawing to demonstrate knowledge of closed-class vocabulary and associated morphosyntactic conventions.

Basic Concepts. The child hears a simple command and points to a color drawing that represents a target adjective or short adjectival phrase.

Sentence Structure. The child hears sentences of varying length and points to a color drawing that represents the meaning of the sentence. Targets include pronouns, prepositional phrases, verb tenses and aspects, subordinate clauses, and other morphosyntactic forms.

Language Production

Expressive Vocabulary Test (EVT; Williams, 1997). The child hears an open-class word, sees a color drawing that represents the meaning of the word, and produces a synonym for the word. For 4-year-olds, the first eight items of the EVT follow the same procedure as the Formulating Labels subtest of the CELF–P.

CELF–P

Recalling Sentences. The child hears and immediately repeats sentences (of increasing length and complexity) within the context of a story, which the examiner reads while the child views color illustrations.

Formulating Labels. The child sees a color drawing, hears a ''what'' question about the drawing, and produces an open-class word (noun or verb) that answers the question.

Word Structure. This subtest uses a cloze procedure with two colored drawings to target morphological forms.

Sequential Intellectual Processing

Kaufman Assessment Battery for Children (K–ABC; Kaufman & Kaufman, 1983)

Hand Movements. The child sees hand gestures and immediately reproduces the sequence manually, demonstrating nonverbal sequential processing and memory abilities.

Number Recall. This task is called *digit span* in other tests. The child hears digits (in sequences of increasing length) and immediately repeats them.

Word Order. This task is similar to the traditional word span task, except that the output is nonverbal. The child hears names of common objects (in sequences of increasing length) and then points to black-and-white representations of the objects in the same order.

Simultaneous Intellectual Processing

K–ABC *Magic Window*. This is a visual gating task that measures temporal–spatial ability. The child is asked to label drawings of common objects that are passed slowly behind a very small window (so that only a small portion of the drawing is visible at a time).

Gestalt Closure. The child sees an incomplete black-and-white line drawing and produces a verbal label for the partially represented object.

Face Recognition. The child sees a color photo (frontal view) of a face for 5 s and then is asked to identify that person in a group photo.

Triangles. This is an adaptation of block design tasks used in other tests. The child sees a picture of a blue-and-yellow pattern and reproduces the pattern using a set of identical rubber triangles that are blue on one side and yellow on the other.